

```

DT 01-OCT-2003 (TReMBLrel. 25, last annotation update)
DE Site-specific recombinase, phage integrase family.
GN OrderedLocustNames=500388;
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCB1_TaxID=70663;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RC MEDLINE=22297686; PubMed=12368813; DOI=10.1038/bt749;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N.L., Melhe B.A.,
RA Clayton R.A., Meyer T., Tsapin A., Scott J., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S.,
RA Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A.,
RA White O., Wolf A.M., Vamathevan J.J., Weidman J.F., Impraim M.,
RA Lee K., Berry K.J., Lee C., Mueller J., Khouri H.M., Gill J.,
RA Ullrich T.R., McDonald L.A., Feldlyum T.V., Smith H.O.,
RA Venter J.C., Nealon K.H., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015487; AAN53471.1; -.
DR TIGR; S00388; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0015074; P:DNA integration; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR InterPro; IPR011010; DNA_birk_join_enz.
DR InterPro; IPR002104; Phage_integrate.
DR Pfam; PF00589; Phage_integrase; 1.
KW Complete proteome.
SQ
SEQUENCE 429 AA; 48452 MW; E6AC2F3ED4E113D CRC64;

Query Match
Best Local Similarity 94.3%; Score 33; DB 2; Length 429;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 294 DSGELKR 300

RESULT 3
Q9PP43 PRELIMINARY; PRT; 643 AA.
AC Q9PP43;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, last annotation update)
DE ABC transport system ATP-binding protein.
GN OrderedLocustNames=Cj0888.3, Cj0888c;
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCB1_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RC MEDLINE=20150912; PubMed=10688204; DOI=10.1038/35001088;
RA Parkhill J., Wren B.W., Mungall K.L., Kelsey J.M., Churcher C.M.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA Jajelski K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rastall M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AL139076; CAB73146.1; -.
DR PIR; H81361; H81361.
DR HSSP; P58301; IUS8.
DR GO; GO:0016020; C:membrane; IEA.

```

```

DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042625; F:ATPase activity; coupled to transmembrane m. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Complete proteome.
SQ
SEQUENCE 643 AA; 73915 MW; F23FCA0DFE0CB0D5 CRC64;

Query Match
Best Local Similarity 94.3%; Score 33; DB 2; Length 643;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 384 DSGELKR 390

RESULT 4
RL32 BRAJA
ID RL32 BRAJA STANDARD; PRT; 60 AA.
AC Q89V07;
DT 25-OCT-2004 (rel. 45, Created)
DT 25-OCT-2004 (rel. 45, last sequence update)
DT 25-OCT-2004 (rel. 45, last annotation update)
DE 50S ribosomal protein L32.
GN Name-rpmf; OrderedLocustNames=br0948;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCB1_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RC MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Ideasa K., Itiguchi M., Kawashima K.,
RA Kohara M., Matsunoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
CC -1- SIMILARITY: Belongs to the ribosomal protein L32P family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; AP005938; BAC46213.1; -.
CC DR HAMAP; MF_00340; -.
CC DR InterPro; IPR002677; Ribosomal_L32P.
CC DR InterPro; IPR005718; Ribosomal_L32b/o.
CC DR Pfam; PF01783; Ribosomal_L32P; 1.
CC DR TIGRPFam; TIGR01031; rpmf_bact; 1.
CC KW Complete proteome; Ribosomal protein.
SQ
SEQUENCE 60 AA; 6979 MW; 5D0F3B445AF136D9 CRC64;

Query Match
Best Local Similarity 91.4%; Score 32; DB 1; Length 60;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 32 DSGELKR 38

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds  
(without alignments)

51.825 Million cell updates/sec

Title: US-10-646-919-37

Perfect score: 35

Sequence: 1 DSGELKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : uniprot\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	199	2	Q8PA66 xanthomonas
2	33	94.3	429	2	Q8EJS4 shewanella
3	33	94.3	643	2	Q9PP43 campylobact
4	32	91.4	60	1	RL32_BRAJA
5	32	91.4	60	1	RL32_RHOFA
6	32	91.4	62	1	RL32_METEX
7	32	91.4	330	2	Q96LH8
8	32	91.4	367	2	Q8NEB2
9	32	91.4	480	2	Q9LSK7
10	32	91.4	495	2	Q683K1
11	32	91.4	495	2	Q8GW96
12	32	91.4	617	2	Q6ZH27
13	32	91.4	617	2	Q6JQU2
14	32	91.4	910	2	Q7RSD3
15	32	91.4	996	2	Q9S112
16	31	88.6	216	2	Q8JDF7
17	31	88.6	223	2	Q89GT6
18	31	88.6	248	2	Q8A522
19	31	88.6	384	1	YAGA_ECOLI
20	31	88.6	534	2	Q883N3
21	31	88.6	612	2	Q9C927
22	31	88.6	737	2	Q9CJS2
23	31	88.6	825	2	Q6FBW4
24	31	88.6	828	2	Q87KJ8
25	31	88.6	1028	1	CARB_THET2
26	31	88.6	1630	2	Q70073
27	31	88.6	3940	2	Q9HRI9
28	31	88.6	3941	2	Q787K6
29	30	85.7	75	2	Q70235
30	30	85.7	111	2	Q8M5F0
31	30	85.7	111	2	Q7G656

32	30	85.7	114	2	Q6UD64	Q6UD64 uncultured
33	30	85.7	165	2	Q9AGB1	Q9AGB1 borrelia bu
34	30	85.7	172	2	Q8TLR8	Q8TLR8 methanosarc
35	30	85.7	172	2	Q8GR55	Q8GR55 borrelia va
36	30	85.7	185	2	Q31116	Q31116 borrelia bu
37	30	85.7	188	2	Q6RU9	Q6RU9 borrelia ga
38	30	85.7	188	2	Q6WQP5	Q6WQP5 borrelia bu
39	30	85.7	189	2	Q6WQP7	Q6WQP7 borrelia bu
40	30	85.7	191	2	Q31120	Q31120 borrelia bu
41	30	85.7	192	2	Q9S3P2	Q9S3P2 borrelia bu
42	30	85.7	197	2	Q8GR57	Q8GR57 borrelia va
43	30	85.7	200	2	Q86HP7	Q86HP7 dictyostell
44	30	85.7	210	2	Q44719	Q44719 borrelia bu
45	30	85.7	210	2	Q45176	Q45176 borrelia ga

## ALIGNMENTS

RESULT 1	Q8PA66	PRELIMINARY;	PRT;	199 AA.
ID	Q8PA66			
AC	Q8PA66			
DT	01-OCT-2002 (TRMBLrel. 22, Created)			
DT	01-OCT-2002 (TRMBLrel. 22, Last sequence update)			
DT	01-OCT-2002 (TRMBLrel. 22, Last annotation update)			
DE	Hypothetical protein XCC1623.			
GN	OrderedLocusNames=XCC1623;			
OS	Xanthomonas campestris (pv. campestris).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;			
OC	Xanthomonadaceae; Xanthomonas.			
OX	NCBI_Taxid=340;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 33913 / NCPPB 528;			
RX	MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;			
RA	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,			
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,			
RA	Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,			
RA	Carrozzini G., Canavan P., Cardozo J., Chamberg F., Ciapina L.P.,			
RA	Ciccarelli R.M.B., Coutinho L.L., Chrsino-Santos J.R., El-Dorry H.,			
RA	Faria J.B., Ferreira A.U.S., Ferreira R.C.C., Ferro W.I.T.,			
RA	Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,			
RA	Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,			
RA	Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,			
RA	Martins E.C., Mendes J., Menck C.F.M., Miyaki C.Y., Moon D.H.,			
RA	Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,			
RA	Pereira H.A., Rossi A., Sena U.A.D., Silva C., de Souza R.F.,			
RA	Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira B.C., Tezza R.I.D.,			
RA	Trindade dos Santos M., Truffi D., Teal S.M., White F.F.,			
RA	Setubal J.C., Kitajima J.P.;			
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing			
RT	host specificities."			
RL	Nature 417:459-463 (2002).			
DR	EMBL; AB012263; AAM40917.1; -.			
KW	Complete proteome.			
SO	SEQUENCE 199 AA; 21899 MW; 6F9D797EC88935B CRC64;			
Query Match	100.0%;	Score 35;	DB 2;	Length 199;
Best Local Similarity	100.0%;	Pred. No. 16;		
Matches	7;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps	0;
QY	1 DSGELKR 7			
DB	115 DSGELKR 121			
RESULT 2	Q8EJS4	PRELIMINARY;	PRT;	429 AA.
ID	Q8EJS4			
AC	Q8EJS4			
DT	01-MAR-2003 (TRMBLrel. 23, Created)			
DT	01-MAR-2003 (TRMBLrel. 23, Last sequence update)			

**THIS PAGE IS BLANK**

Db 763 DSGELK 768

Search completed: July 20, 2005, 13:56:05  
Job time : 15.5 secs

---

A/Map position: 1  
C:Superfamily: Arabidopsis thaliana hypothetical protein At2g41080

Query Match 85.7%; Score 30; DB 2; Length 665;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELK 7  
|||||  
DB 239 DSGELNR 245

## RESULT 12

S48394  
probable membrane protein YII140w - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae  
C>Date: 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 09-Jul-2004  
C/Accession: S48394; S50276  
R:Churcher, C.

submitted to the EMBL Data Library, September 1994  
A/Reference number: S48310  
A/Accession: S48394

A/Molecule type: DNA  
A/Accession: 1-823 <CHU>

A/Residues: 1-823 <CHU>  
A/Cross-references: UNIPROT:P38928; GB:Z47047; EMBL:Z38059; NID:G603997; PID:G763206; MI  
R:Torrey, L.E.; Gibbs, P.E.M.; Nelson, J.; Lawrence, C.W.  
Yeast 10, 1503-1509, 1994

A/Title: Cloning and sequence of REV7, a gene whose function is required for DNA damage-  
A/Reference number: S50275; MUID:95176709; PMID:7871890  
A/Accession: S50276

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA  
A/Residues: 80-823 <TOR>

A/Cross-references: EMBL:U07228; NID:G460247; PID:AAA67919.1; PID:G460249

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994

C:Genetics: SGD:SRO4

A/Map position: 9L  
C:Superfamily: Saccharomyces cerevisiae probable membrane protein YII140w

C/Keywords: transmembrane protein  
F:5-22/Domain: transmembrane #status predicted <TM1>  
F:511-527/Domain: transmembrane #status predicted <TM2>

Query Match 85.7%; Score 30; DB 2; Length 823;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6  
|||||  
DB 198 DSGELK 203

## RESULT 13

A57238  
iron-responsive element-binding protein 2, hepatic - rat

M/Alternate names: iron regulatory protein 2  
C:Species: Rattus norvegicus (Norway rat)

C>Date: 08-Dec-1995 #sequence\_revision 08-Dec-1995 #text\_change 16-Jul-1999  
C/Accession: A57238

R:Guo, B.; Brown, F.M.; Phillips, J.D.; Yu, Y.; Leibold, E.A.  
J. Biol. Chem. 270, 16529-16535, 1995

A/Title: Characterization and expression of iron regulatory protein 2 (IRP2). Presence of  
A/Reference number: A57238; MUID:95348066; PMID:7622457

A/Accession: A57238  
A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA  
A/Residues: 1-963 <GUO>

A/Cross-references: GB:U20181; NID:G897582; PID:AAA79927.1; PID:G897583

A/Experimental source: strain Sprague-Dawley; tissue type adult liver  
C:Genetics:  
A/Map position: 1

A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1254 <STO>  
A/Cross-references: UNIPROT:Q9SR21; GB:AE005172; NID:G6056403; PID:AAF02867.1; GSPDB:G  
C:Genetics:  
A/Map position: 1

C/Keywords: 4Fe-4S; iron-sulfur protein; liver; metalloprotein; RNA binding  
F:555-698/Domain: RNA binding #status predicted <RNA>  
F:512,578,581/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 85.7%; Score 30; DB 2; Length 963;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELK 7  
|||||  
DB 180 DSGELSR 186

## RESULT 14

D97001  
probable membrane protein [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum  
C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: D97001

R:Nolling, U.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J. Bacteriol. 183, 4823-4838, 2001

A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A/Reference number: A96900; MUID:21359325; PMID:21359325

A/Accession: D97001  
A/Status: preliminary

A/Molecule type: DNA  
A/Residues: 1-1116 <KOR>

A/Cross-references: UNIPROT:Q97KJ3; GB:AE001437; PID:AAK78799.1; PID:G15023714; GSPDB:G  
A/Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics: CAC0823

Query Match 85.7%; Score 30; DB 2; Length 1116;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6  
|||||  
DB 865 DSGELK 870

## RESULT 15

E86160  
hypothetical protein F22D16.1 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C/Accession: E86160

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

Ker, M.; Wu, D.; Yu, G.; Frazer, C.W.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: E86160  
A/Status: preliminary

A/Molecule type: DNA  
A/Residues: 1-1254 <STO>

A/Cross-references: UNIPROT:Q9SR21; GB:AE005172; NID:G6056403; PID:AAF02867.1; GSPDB:G  
C:Genetics:  
A/Map position: 1

Query Match 85.7%; Score 30; DB 2; Length 1254;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6  
|||||

A:Cross-references: UNIPROT:Q44976; GB:U25413; NID:g9495735; PIDN:AAA22956.1; PID:g9495736  
C:Gene: ospC  
A:Gene: ospC  
C:Superfamily: Lyme disease epirochete surface protein C

Query Match 85.7%; Score 30; DB 2; Length 211;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSGELKR 6  
|||  
Db 117 DSGELK 122

## RESULT 8

B96759 protein serine carboxypeptidase T18K17.4 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 16-Aug-2004

C:Accession: B96759

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A66141; MUID:21016719; PMID:11130712

A:Accession: B96759

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-438 <STO>

A:Cross-references: UNIPROT:Q9CAU2; GB:AE005173; NID:g6598854; PIDN:AAF18708.1; GSPDB:GN

C:Gene: T18K17.4

A:Map position: 1

C:Superfamily: Serine carboxypeptidase

Query Match 85.7%; Score 30; DB 2; Length 438;  
Best Local Similarity 85.7%; Pred. No. 96;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DSGELKR 7  
|||  
Db 151 DSGELK 157

## RESULT 9

A96759 protein serine carboxypeptidase T18K17.5 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 16-Aug-2004

C:Accession: A96759

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A66141; MUID:21016719; PMID:11130712

A:Accession: A96759

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <STO>

A:Cross-references: UNIPROT:Q9CAU1; GB:AE005173; NID:g6598855; PIDN:AAF18709.1; GSPDB:GN

C:Gene: T18K17.5

A:Gene: T18K17.5  
A:Map position: 1  
C:Superfamily: Serine carboxypeptidase

Query Match 85.7%; Score 30; DB 2; Length 441;  
Best Local Similarity 85.7%; Pred. No. 96;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DSGELKR 7  
|||  
Db 152 DSGELK 158

## RESULT 10

C96759 protein serine carboxypeptidase T18K17.3 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 16-Aug-2004

C:Accession: C96759

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A66141; MUID:21016719; PMID:11130712

A:Accession: C96759

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <STO>

A:Cross-references: UNIPROT:Q9CAU3; GB:AE005173; NID:g6598853; PIDN:AAF18707.1; GSPDB:GN

C:Gene: T18K17.3

A:Map position: 1

C:Superfamily: Serine carboxypeptidase

Query Match 85.7%; Score 30; DB 2; Length 441;  
Best Local Similarity 85.7%; Pred. No. 96;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DSGELKR 7  
|||  
Db 152 DSGELK 158

## RESULT 11

F66181 protein F13M7.17 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: F66181

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A66141; MUID:21016719; PMID:11130712

A:Accession: F66181

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-665 <STO>

A:Cross-references: UNIPROT:Q9MAT2; GB:AE005172; NID:g7211995; PIDN:AAF40466.1; GSPDB:GN

C:Gene: F13M7.17

Best Local Similarity 85.7%; Pred. No. 80;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 7  
|||||

DB 281 DNGELKR 287

## RESULT 3

C64752  
outer surface protein C precursor - *Escherichia coli* (strain K-12)

C/Species: *Escherichia coli*

C/Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004

C/Accession: C64752

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shaoh, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of *Escherichia coli* K-12.

A/Reference number: A64720; PMID:9742617; PMID:9278503

A/Accession: C64752

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-384 <BLAT>

A/Cross-references: UNIPROT:P37007; GB:AE000134; GB:U00096; NID:g1786454; PIDN:AACT3370.

A/Experimental source: strain K-12, substrain MGL55

C/Genetics:

A/Gene: yagA

## Query Match

Best Local Similarity 88.6%; Score 31; DB 2; Length 384;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 7  
|||||

DB 266 DSGELQR 272

## RESULT 4

I40274  
outer surface protein C precursor - *Borrelia* sp.

C/Species: *Borrelia* sp.

C/Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 26-May-2000

C/Accession: I40274

R/Fukunaga, M.; Hamase, A.

J. Clin. Microbiol. 33, 2415-2420, 1995

A/Title: Outer surface protein C gene sequence analysis of *Borrelia burgdorferi* sensu la

A/Reference number: I40269; PMID:96025162; PMID:7494039

A/Accession: I40274

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-210 <RES>

A/Cross-references: GB:D49381; NID:g1041109; PIDN:BAA08379.1; PID:g1041110

C/Superfamily: Lyme disease spirochete surface protein C

## Query Match

Best Local Similarity 85.7%; Score 30; DB 2; Length 210;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6  
|||||

DB 114 DSGELK 119

## RESULT 5

I40284  
outer surface protein C precursor - *Borrelia* sp.

C/Species: *Borrelia* sp.

C/Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 26-May-2000

C/Accession: I40284; I40283

R/Fukunaga, M.; Hamase, A.

J. Clin. Microbiol. 33, 2415-2420, 1995

A/Title: Outer surface protein C gene sequence analysis of *Borrelia burgdorferi* sensu la

A/Reference number: I40269; PMID:96025162; PMID:7494039

A/Accession: I40284  
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-210 <RES>

A/Cross-references: GB:D49508; NID:g707103; PIDN:BAA08468.1; PID:g769695

A/Accession: I40283

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-210 <RE2>

A/Cross-references: GB:D49507; NID:g707102; PID:g769694

C/Superfamily: Lyme disease spirochete surface protein C

Query Match

Best Local Similarity 85.7%; Score 30; DB 2; Length 210;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6  
|||||

DB 114 DSGELK 119

## RESULT 6

I40144  
outer surface protein C precursor - Lyme disease spirochete

C/Species: *Borrelia burgdorferi* (Lyme disease spirochete)

C/Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 09-Jul-2004

C/Accession: I40144; S70282

R/Stevenson, B.; Barthold, S.W.

FEMS Microbiol. Lett. 124, 367-372, 1994

A/Title: Expression and sequence of outer surface protein C among North American isolate

A/Reference number: I40143; PMID:95154673; PMID:7851744

A/Accession: I40144

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-210 <RES>

A/Cross-references: UNIPROT:Q44719; EMBL:U04281; NID:g434663; PIDN:AACT3297.1; PID:g4346

R/Lively, I.; Gibbs, C.P.; Schuster, R.; Dörner, F.

Mol. Microbiol. 18, 257-269, 1995

A/Title: Evidence for lateral transfer and recombination in OspC variation in Lyme disea

A/Reference number: S70255; PMID:96296448; PMID:8709845

A/Accession: S70282

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 19-210 <LIV>

A/Cross-references: EMBL:I42893; NID:g858721; PIDN:AAB37001.1; PID:g1695218

A/Experimental source: strain 297

C/Genetics:

C/Superfamily: Lyme disease spirochete surface protein C

Query Match

Best Local Similarity 85.7%; Score 30; DB 2; Length 210;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6  
|||||

DB 116 DSGELK 121

## RESULT 7

I40268  
outer surface protein C precursor - Lyme disease spirochete

C/Species: *Borrelia burgdorferi* (Lyme disease spirochete)

C/Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 09-Jul-2004

C/Accession: I40268

R/Margolis, N.; Hogan, D.; Cieplak, W.

Gene 143, 105-110, 1994

A/Title: Homology between *Borrelia burgdorferi* OspC and members of the family of *Borrelli*

A/Reference number: I40268; PMID:94259285; PMID:8200524

A/Accession: I40268

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-211 <RES>

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds  
(without alignments)  
46.449 Million cell updates/sec

Title: US-10-646-919-37

Perfect score: 35  
Sequence: 1 DSGELKR 7Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	643	2	ABCB transport syst
2	32	91.4	996	2	probable AAA-type
3	31	88.6	384	2	Yaga protein - Bsc
4	30	85.7	210	2	outer surface prot
5	30	85.7	210	2	outer surface prot
6	30	85.7	210	2	outer surface prot
7	30	85.7	211	2	outer surface prot
8	30	85.7	438	2	protein serine car
9	30	85.7	441	2	protein serine car
10	30	85.7	441	2	protein serine car
11	30	85.7	665	2	protein serine car
12	30	85.7	823	2	protein F13M7.17 (
13	30	85.7	963	2	probable membrane
14	30	85.7	1116	2	iron-responsive el
15	30	85.7	1254	2	probable membrane
16	30	85.7	1290	2	hypothetical prote
17	30	85.7	1613	2	dextranucrase (EC
18	30	85.7	1871	2	hypothetical prote
19	30	85.7	3705	2	probable heat choc
20	29	82.9	60	2	probable autotrans
21	29	82.9	217	2	ribosomal protein
22	29	82.9	235	2	hypothetical prote
23	29	82.9	349	2	probable biotin-la
24	29	82.9	357	2	TYA protein - yeas
25	29	82.9	379	2	probable methylate
26	29	82.9	405	2	citrate (si)-synth
27	29	82.9	430	2	hypothetical prote
28	29	82.9	430	2	probable serine ca
29	29	82.9	430	2	probable valine-py
29	29	82.9	430	2	valine-pyruvate am

30	29	82.9	433	2	A84619	probable serine ca
31	29	82.9	438	2	S70231	TYA protein - yeas
32	29	82.9	438	2	S70232	TYA protein - yeas
33	29	82.9	438	2	S19347	TYA protein - yeas
34	29	82.9	438	2	S61574	TYA protein - yeas
35	29	82.9	438	2	S45397	TYA protein - yeas
36	29	82.9	438	2	S67320	TYA protein - yeas
37	29	82.9	438	2	S69968	TYA protein - yeas
38	29	82.9	438	2	S69965	TYA protein - yeas
39	29	82.9	438	2	S61589	TYA protein - yeas
40	29	82.9	438	2	F69440	conserved hypothet
41	29	82.9	461	2	E75304	hypothetical prote
42	29	82.9	472	2	G83537	aromatic amino aci
43	29	82.9	478	2	S69974	TYA protein - yeas
44	29	82.9	490	2	S76169	hypothetical prote
45	29	82.9	497	2	G95398	probable ABC trans

## ALIGNMENTS

## RESULT 1

H81361

ABC transport system ATP-binding protein Cj0888c [imported] - Campylobacter jejuni (str.

Cj:Species: Campylobacter jejuni  
Cj:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #ext\_change 16-Aug-2004Cj:Accession: H81361  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kellay, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell

Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyA:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: H81361A:Status: Preliminary  
A:Molecule type: DNAA:Residues: 1-643 <PAR>  
A:Cross-references: UNIPROT:Q9PP43; GB:AL139076; GB:AL111168; NID:96968128; PIDN:CAB731-A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:A:Gene: Cj0888c  
C:Superfamily: ATP-binding cassette homology

Query Match 94.3%; Score 33; DB 2; Length 643;  
Best Local Similarity 85.7%; Pred. No. 30;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7

DB 384 DSGELKR 390

## RESULT 2

D84561

probable AAA-type ATPase [imported] - Arabidopsis thaliana  
Cj:Species: Arabidopsis thaliana (mouse-ear cross)Cj:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #ext\_change 09-Jul-2004  
R:Jin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentley, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanNaken, S.E.; Umayam, L.; Tallon, L

eues, D.; Nierman, W.C.; White, O.; Eilen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, C

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197A:Accession: D84561  
A:Status: preliminaryA:Molecule type: DNA  
A:Residues: 1-996 <STO>

A:Cross-references: UNIPROT:Q9SI12; GB:AE002093; NID:94874284; PIDN:AAD31347.1; GSPDB:G

A:Gene: At2g18190  
A:Map position: 2

Query Match 91.4%; Score 32; DB 2; Length 996;



```
RESULT 2
US-10-282-122A-54450
; Sequence 54450, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyeh, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR FILING DATE: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR FILING DATE: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR FILING DATE: 2001-02-09
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54450
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
US-10-282-122A-54450

Query Match          94.3%; Score 33; DB 15; Length 643;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELKR 7
DB      384 DSGELKR 390

RESULT 3
US-09-764-891-3622
; Sequence 3622, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 3622
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (54)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (77)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3622

Query Match          91.4%; Score 32; DB 10; Length 80;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELKR 7
DB      47 DSGELKR 53

RESULT 4
US-10-424-599-234622
; Sequence 234622, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234622
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53892C.1.pep
US-10-424-599-234622

Query Match          91.4%; Score 32; DB 15; Length 287;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELKR 7
DB      122 DSGDLKR 128

RESULT 5
US-10-425-114-44704
; Sequence 44704, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven B
; APPLICANT: Tabaska, Jack B
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:52:31 ; Search time 71.333 Seconds  
(without alignments)  
38.093 Million cell updates/sec

Title: US-10-646-919-37

Perfect score: 35  
Sequence: 1 DSGELKR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388189149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubppaa/US11\_PUBCOMB.pep:\*  
20: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep:\*  
21: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
22: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	100.0	7	US-10-646-919-37	Sequence 37, Appl
2	33	94.3	643	US-10-282-122A-54450	Sequence 54450, A
3	32	91.4	80	US-09-764-891-3622	Sequence 3622, Ap
4	32	91.4	287	US-10-424-599-234622	Sequence 234622, A
5	32	91.4	292	US-10-425-114-44704	Sequence 44704, A
6	32	91.4	339	US-10-188-248-82	Sequence 82, Appl
7	32	91.4	340	US-10-188-248-84	Sequence 84, Appl
8	32	91.4	480	US-10-225-066A-496	Sequence 496, Appl
9	32	91.4	480	US-10-374-780A-2902	Sequence 2902, Ap
10	32	91.4	480	US-10-669-824-60	Sequence 60, Appl
11	32	91.4	480	US-10-870-198-60	Sequence 60, Appl

12	32	91.4	522	15	US-10-425-114-56109	Sequence 56109, A
13	31	88.6	257	15	US-10-424-599-163059	Sequence 163059, A
14	30	85.7	99	16	US-10-767-701-31966	Sequence 31966, A
15	30	85.7	111	16	US-10-437-963-143516	Sequence 143516, A
16	30	85.7	120	16	US-10-437-963-145261	Sequence 145261, A
17	30	85.7	208	16	US-10-437-963-133203	Sequence 133203, A
18	30	85.7	222	16	US-10-437-963-133982	Sequence 133982, A
19	30	85.7	270	16	US-10-767-701-38282	Sequence 38282, A
20	30	85.7	283	15	US-10-369-493-13574	Sequence 13574, A
21	30	85.7	300	9	US-09-738-626-4711	Sequence 4711, Ap
22	30	85.7	317	16	US-10-425-115-244452	Sequence 244452, A
23	30	85.7	341	16	US-10-437-963-131112	Sequence 131112, A
24	30	85.7	415	15	US-10-369-493-7947	Sequence 7947, Ap
25	30	85.7	426	15	US-10-425-114-53923	Sequence 53923, A
26	30	85.7	463	17	US-10-470-048B-577	Sequence 577, Appl
27	30	85.7	465	15	US-10-282-122A-70724	Sequence 70724, A
28	30	85.7	465	18	US-10-724-972A-5599	Sequence 5599, Ap
29	30	85.7	499	16	US-10-767-701-46391	Sequence 46391, A
30	30	85.7	511	16	US-10-437-963-125840	Sequence 125840, A
31	30	85.7	526	15	US-10-425-114-62867	Sequence 62867, A
32	30	85.7	583	16	US-10-437-963-114326	Sequence 114326, A
33	30	85.7	584	16	US-10-425-115-348909	Sequence 348909, A
34	30	85.7	594	16	US-10-437-963-137342	Sequence 137342, A
35	30	85.7	779	16	US-10-437-963-115617	Sequence 115617, A
36	30	85.7	963	9	US-09-924-396B-20	Sequence 20, Appl
37	30	85.7	3705	15	US-10-282-122A-77944	Sequence 77944, A
38	29	82.9	7	18	US-10-646-919-53	Sequence 53, Appl
39	29	82.9	82	15	US-10-424-599-180374	Sequence 180374, A
40	29	82.9	103	16	US-10-425-115-258364	Sequence 258364, A
41	29	82.9	105	16	US-10-425-115-225746	Sequence 225746, A
42	29	82.9	106	17	US-10-732-923-12434	Sequence 12434, A
43	29	82.9	159	15	US-10-424-599-227781	Sequence 227781, A
44	29	170	15	15	US-10-424-599-171900	Sequence 171900, A
45	29	82.9	187	16	US-10-437-963-125509	Sequence 125509, A

#### ALIGNMENTS

RESULT 1  
US-10-646-919-37  
; Sequence 37, Application US/10646919  
; Publication No. US20050148075A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbas, Carlos F.  
; TITLE OF INVENTION: Zinc-finger binding domains for GNN  
; FILE REFERENCE: TSRI 645.2C1  
; CURRENT APPLICATION NUMBER: US/10/646,919  
; CURRENT FILING DATE: 2003-08-21  
; PRIOR APPLICATION NUMBER: US 09/494,190  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: PCT/EP99/07742  
; PRIOR FILING DATE: 1999-10-14  
; PRIOR APPLICATION NUMBER: US 09/173,941  
; PRIOR FILING DATE: 1998-10-16  
; NUMBER OF SEQ ID NOS: 129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 37  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthesized  
US-10-646-919-37

Query Match 100.0%; Score 35; DB 18; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7  
| | | | |  
DB 1 DSGELKR 7

**THIS PAGE IS BLANK**

Thu Jul 21 08:53:08 2005

Qy	1	DSGEIKR	7
Db	200	DSGEIKR	206

Search completed: July 20, 2005, 13:54:28  
Job time : 22.6667 secs

```

; LENGTH: 224
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (76)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-482-273-174

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 224;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELKR 7
Db 188 DLGELKR 194

RESULT 12
US-09-902-540-10834
; Sequence 10834, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10834
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-10834

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 300;
Best Local Similarity 71.4%; Pred. No. 1.9e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
Db 194 DLGELKR 200

RESULT 13
US-09-482-273-105
; Sequence 105, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 105
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```

US-09-482-273-105

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 302;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELKR 7
Db 188 DLGELKR 194

RESULT 14
US-09-252-991A-31133
; Sequence 31133, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31133
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31133

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SGEIKR 7
Db 169 SGEIKR 174

RESULT 15
US-09-538-092-21
; Sequence 21, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Iolita
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 21
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YBL101W-A
US-09-538-092-21

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 438;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
RESULT 7
; US-09-710-279-658
; Sequence 658, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; US-09-710-279-658
```

```
Query Match      85.7%; Score 30; DB 4; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 DSGELK 6
        |||||
Db      401 DSGELK 406
```

```
RESULT 8
; US-09-173-941-53
; Sequence 53, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 53
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; US-09-173-941-53
```

```
Query Match      82.9%; Score 29; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 SGE LK R 7
        |||||
Db      2 SGE LK R 7
```

```
RESULT 9
; US-09-494-190-53
; Sequence 53, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
```

```
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 53
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
; US-09-494-190-53
```

```
Query Match      82.9%; Score 29; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 SGE LK R 7
        |||||
Db      2 SGE LK R 7
```

```
RESULT 10
; US-09-270-767-60792
; Sequence 60792, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 60792
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
; US-09-270-767-60792
```

```
Query Match      82.9%; Score 29; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 SGE LK R 7
        |||||
Db      60 SGE LK R 65
```

```
RESULT 11
; US-09-482-273-174
; Sequence 174, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 174
```

FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:phage display  
OTHER INFORMATION: selected and mutagenized  
US-09-494-190-37

Query Match 100.0%; Score 35; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.1e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 7  
| | | | |  
DB 1 DSGELK 7

RESULT 3  
US-09-248-796A-18533  
Sequence 18533, Application US/09248796A

PATENT INFORMATION:  
PATENT NO. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstein et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A

PRIOR FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725

PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409

NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 18533  
LENGTH: 612  
TYPE: PRT  
ORGANISM: Candida albicans

Query Match 88.6%; Score 31; DB 4; Length 612;  
Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELK 7  
| | | | |  
DB 580 DSGELK 586

RESULT 4  
US-08-671-548C-12  
Sequence 12, Application US/08671548C

PATENT NO. 6486130  
GENERAL INFORMATION:

APPLICANT: LIVEY, Ian  
APPLICANT: CROWE, Brian  
TITLE OF INVENTION: IMMUNOGENIC FORMULATION OF OSCP ANTIGEN VACCINE FOR THE PREVENTIO  
TITLE OF INVENTION: TREATMENT OF LYME DISEASE AND RECOMBINANT METHODS FOR THE PREPAR  
FILE REFERENCE: 37974-0023  
CURRENT APPLICATION NUMBER: US/08/671,548C

PRIOR FILING DATE: 2001-08-30  
PRIOR APPLICATION NUMBER: US 08/284,667

PRIOR FILING DATE: 1994-08-19  
PRIOR APPLICATION NUMBER: 08/053,863

PRIOR FILING DATE: 1993-04-29  
PRIOR APPLICATION NUMBER: PCT/EP94/01365

NUMBER OF SEQ ID NOS: 78  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 12  
LENGTH: 192  
TYPE: PRT  
ORGANISM: Borrelia burgdorferi  
US-08-671-548C-12

Query Match 85.7%; Score 30; DB 4; Length 192;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6  
| | | | |  
DB 98 DSGELK 103

RESULT 5  
US-09-345-236B-43  
Sequence 43, Application US/09345236B

PATENT NO. 6521454  
GENERAL INFORMATION:  
APPLICANT: Becnel, James J.  
APPLICANT: Tokuo, Fukuda  
APPLICANT: Moser, Bettina  
APPLICANT: Cockburn, Andrew  
APPLICANT: White, Susan E.  
APPLICANT: Undeen, Albert H.

TITLE OF INVENTION: No. 6521454e1 Baculoviruses, Insecticidal  
TITLE OF INVENTION: Compositions, and Methods for Control of Invertebrates  
FILE REFERENCE: 21042.0004  
CURRENT APPLICATION NUMBER: US/09/345,236B

PRIOR FILING DATE: 1999-06-30  
NUMBER OF SEQ ID NOS: 148  
SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 43  
LENGTH: 433  
TYPE: PRT  
ORGANISM: mosquito baculovirus  
US-09-345-236B-43

Query Match 85.7%; Score 30; DB 4; Length 433;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELK 7  
| | | | |  
DB 130 DSGELK 136

RESULT 6  
US-09-134-001C-3856  
Sequence 3856, Application US/09134001C

PATENT NO. 6380370  
GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C

PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3856  
LENGTH: 465  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3856

Query Match 85.7%; Score 30; DB 3; Length 465;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6  
| | | | |  
DB 401 DSGELK 406

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds  
(without alignments)  
24.117 Million cell updates/sec

Title: US-10-646-919-37

Perfect score: 35

Sequence: 1 DSGELKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysts of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	7	3 US-09-173-941-37	Sequence 37, App1
2	35	100.0	7	4 US-09-494-190-37	Sequence 37, App1
3	31	88.6	612	4 US-09-248-796A-18533	Sequence 18533, A
4	30	85.7	192	4 US-08-671-548C-12	Sequence 12, App1
5	30	85.7	433	4 US-09-345-236B-43	Sequence 43, App1
6	30	85.7	465	3 US-09-134-001C-3856	Sequence 3856, App
7	30	85.7	472	4 US-09-710-279-658	Sequence 658, App
8	29	82.9	7	3 US-09-173-941-53	Sequence 53, App1
9	29	82.9	7	4 US-09-494-190-53	Sequence 53, App1
10	29	82.9	66	4 US-09-270-767-60792	Sequence 60792, A
11	29	82.9	224	4 US-09-482-273-174	Sequence 174, App
12	29	82.9	300	4 US-09-902-540-10834	Sequence 10834, A
13	29	82.9	302	4 US-09-482-273-105	Sequence 105, App
14	29	82.9	365	4 US-09-252-991A-31133	Sequence 31133, A
15	29	82.9	438	4 US-09-538-092-21	Sequence 21, App1
16	29	82.9	438	4 US-09-538-092-62	Sequence 62, App1
17	29	82.9	475	4 US-09-252-991A-29697	Sequence 29697, A
18	29	82.9	640	4 US-09-907-794A-292	Sequence 292, App
19	29	82.9	640	4 US-09-905-125A-292	Sequence 292, App
20	29	82.9	640	4 US-09-902-775A-292	Sequence 292, App
21	29	82.9	640	4 US-09-906-700-292	Sequence 292, App
22	29	82.9	640	4 US-09-903-603A-292	Sequence 292, App
23	29	82.9	640	4 US-09-904-920A-292	Sequence 292, App
24	29	82.9	640	4 US-09-909-064-292	Sequence 292, App
25	29	82.9	640	4 US-09-905-381A-292	Sequence 292, App
26	29	82.9	640	4 US-09-906-618-292	Sequence 292, App
27	29	82.9	718	4 US-09-252-991A-25696	Sequence 25696, A

28	29	82.9	830	4 US-09-252-991A-22004	Sequence 22004, A
29	29	82.9	1191	4 US-09-921-099A-2	Sequence 2, App1
30	29	82.9	1191	4 US-09-921-099A-4	Sequence 4, App1
31	28	80.0	7	3 US-09-173-941-36	Sequence 36, App1
32	28	80.0	7	4 US-09-494-190-36	Sequence 64, App1
33	28	80.0	43	2 US-08-488-161-64	Sequence 64, App1
34	28	80.0	43	3 US-09-273-685-64	Sequence 64, App1
35	28	80.0	43	5 PCT-US95-11934-64	Sequence 64, App1
36	28	80.0	139	4 US-09-402-016A-2	Sequence 2, App1
37	28	80.0	181	4 US-09-489-039A-8288	Sequence 8288, App
38	28	80.0	201	4 US-09-786-256C-5	Sequence 5, App1
39	28	80.0	206	4 US-09-270-767-36554	Sequence 36554, A
40	28	80.0	206	4 US-09-270-767-51771	Sequence 51771, A
41	28	80.0	311	4 US-09-902-540-11886	Sequence 11886, A
42	28	80.0	332	4 US-09-902-540-9978	Sequence 9978, App
43	28	80.0	358	4 US-09-784-508-4	Sequence 4, App1
44	28	80.0	436	3 US-08-669-378-8	Sequence 8, App1
45	28	80.0	465	4 US-09-252-991A-21223	Sequence 21223, A

#### ALIGNMENTS

```
RESULT 1
US-09-173-941-37
; Sequence 37, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
US-09-173-941-37

Query Match          100.0%; Score 35; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4; 1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELKR 7
Db      1 DSGELKR 7

RESULT 2
US-09-494-190-37
; Sequence 37, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
```



PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231342P.  
 PR 08-SEP-2000; 2000US-0231343P.  
 PR 08-SEP-2000; 2000US-0231444P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234977P.  
 PR 25-SEP-2000; 2000US-0234988P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236377P.  
 PR 29-SEP-2000; 2000US-0236386P.  
 PR 29-SEP-2000; 2000US-0236399P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241285P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246539P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.

PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 PR XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-465570/50.  
 DR N-PSDB; AAL00934.  
 PT XX  
 PT Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.  
 PS Claim 11; SEQ ID NO 3622; 1297bp + Sequence Listing; English.  
 PS XX  
 CC The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a protein of the invention  
 CC XX  
 SQ Sequence 80 AA;  
 Qy 1 DSGELKR 7  
 Db 47 DSGELKR 53  
 RESULT 6  
 ID ABB95674 standard; protein; 80 AA.  
 XX ABB95674;  
 AC 21-JUN-2002 (first entry)  
 DT Human testicular antigen SEQ ID NO: 1058.  
 DE Human testicular antigen; testes; cancer; metastasis; immune disorder;  
 XX reproductive system disorder; urinary system disorder; gene therapy;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 XX gastrointestinal disease; infection; cystostatic.  
 OS Homo sapiens.  
 XX MO200155317-A2.  
 PN 02-AUG-2001.  
 XX  
 PD  
 XX

RESULT 4  
ABU26526  
ID ABU26526 standard; protein; 643 AA.  
XX  
AC ABU26526;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by prokaryotic essential gene #12053.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Campylobacter jejuni.  
XX  
PN WO20027183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang J, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW,  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR MPI: 2003-029926/02.  
DR N-PSDB; ACA30396.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 54450; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing an antibody capable of specifically binding  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 643 AA;  
XX  
Query Match 94.3%; Score 33; DB 6; Length 643;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DSGELKR 7  
DB 384 DSGELKR 390  
XX  
RESULT 5  
AAM94964  
ID AAM94964 standard; protein; 80 AA.  
XX  
AC AAM94964;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen SEQ ID NO: 3622.  
XX  
DE Human reproductive system related antigen; reproductive system disorder;  
XX  
KW Human; reproductive system related antigen; reproductive system disorder;  
XX  
OS Homo sapiens.  
XX  
PN WO200155320-A2.  
XX  
PD 02-AUG-2001.  
XX  
PI 17-JAN-2001; 2001WO-US001339.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225265P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225368P.  
PR 14-AUG-2000; 2000US-0225370P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225577P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226379P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-SEP-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.

achieved through specific contacts from side chains of amino acid residues in the alpha helix. Each zinc finger can recognise a substrate of 3 bp in target DNA. Covalent linkage of multiple zinc finger domains allows the recognition of extended contiguous asymmetric DNA sequences. For example, a synthetic poly(dedcyl) protein containing six zinc finger domains can recognise an 18 bp sequence, and such proteins are potentially highly gene-specific. The novel nucleotide-binding zinc finger proteins may therefore be used in the development of artificial gene-specific transcriptional regulators. Such transcriptional switches may be used to regulate the expression of oncogenes such as *c-myc*, overexpression of which is involved in malignant transformation. The proteins are therefore useful in the treatment of cancers, and may also be used to activate genes involved in fighting diseases, and to treat viral infections by inhibiting the synthesis of viral gene products. They may be used in DNA-based diagnostic applications. The proteins may also be used in producing functional gene knockout or activation in heterozygous transgenic animals. Proteins of the invention can discriminate between sequences which have a single base difference. This is manifested in a >100-fold decrease in affinity for the variant sequence. Gene activation and repression can be achieved by targeting within the gene transcript, suggesting that information obtained from expressed sequence tags may be sufficient for the construction of gene switches. Sequences AAG2876-802869 represent zinc finger alpha helix phage library peptides disclosed in the invention

Query Match	100.0%	Score 35;	DB 3;	Length 7;
Best Local Similarity	100.0%	Pred. No. 1.Be+06;		
Matches 7; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 DSGELKR 7  
|||  
Db 1 DSGELKR 7

RESULT 2  
AD61986  
ID AD61986 standard; peptide; 7 AA.

AC	ADF61986;
XX	
DT	12-FEB-2004 (first entry)

DE zinc finger binding region #37.

KM zinc finger-nucleotide binding polypeptide; expression regulation;  
KM zinc finger binding region.

OS Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

PF 28-JAN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.

XX  
9

2000

XX	000334/75
DD	

XX  
DE

PT (GNN) n-3', comprises exposing nucleotide sequence to composition

PT polypeptide.  
VV

PS Disclosure; SEQ ID NO 37; 46pp; English.  
 YY

CC The invention relates to an expression of nucleotide sequence that  
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by  
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by  
CC exposing nucleotide sequence to composition comprising isolated and  
CC purified zinc finger-nucleotide binding polypeptide that contains a  
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-  
CC Arg. The invention is used in the regulation of the expression of the  
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present  
CC sequence represents the amino acid sequence of a zinc finger binding  
CC region.  
XX  
CC  
XX Sequence 7 AA;  
SQ

Query Match	100.0%	Score 35;	DB 7;	Length 7;
Best Local Similarity	100.0%	Pred. No. 1.8e+06;		
Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 DSGELKR 7  
| | | | |  
Db 1 DSGELKR 7

RESULT 3  
ADJ98428  
ID ADJ98428 standard; peptide; 7 AA.

DT	06-MAY-2004	(first entry)
XX		
DE	zinc finger DNA binding peptide #99.	

KM library; multimeric DNA binding polypeptide;  
KM zinc finger DNA binding peptide; gene expression silencing;  
KM gene expression enhancement.

OS Unidentified.

PN WO2003066828-A2.

PD 14-AUG-2003.

PF 07-FEB-2003; 2003WO-US003705.

PR 07-FEB-2002; 2002US-0354981P.

PA (SCRI ) SCRIPPS RES INST.

PI Barbas CF, Blancafort P;

DR WPI; 2003-731499/69.

PT New zinc finger library of multimeric DNA binding polypeptides, useful

PT factor, and enhancing or silencing target gene expression.

PS Disclosure; Fig 20; 64pp; English.

The invention comprises a library of multimeric DNA binding polypeptides

polypeptides of the invention are useful for sterically occluding the

target gene expression. The present amino acid sequence represents a zinc

3 5

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Best Local Similarity 100.0%; Pred. No. 1.8e+06;

1 RECEIVED 7

1 DECEMBER 7

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      July 20, 2005, 13:14:30 ; Search time 82.5 Seconds
              (without alignments)
              32.816 Million cell updates/sec
```

Title: US-10-646-919-37

Perfect score: 35  
Sequence: 1 DSGELKR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :
A: Geneseqp16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	100.0	7	3	AAB02896	Aab02896 Nucleotide
2	35	100.0	7	7	ADf61986	Adf61986 Zinc finger
3	35	100.0	7	7	ADj98428	Adj98428 Zinc finger
4	33	94.3	643	6	ABU26526	Abu26526 Protein e
5	32	91.4	80	4	AAM94964	Aam94964 Human rep
6	32	91.4	80	4	ABB95674	Abb95674 Human tes
7	32	91.4	104	4	ABM65048	Aum65048 Propionib
8	32	91.4	104	6	ABM61567	Abm61567 Propionib
9	32	91.4	339	7	ADc13603	Adc13603 Human NOV
10	32	91.4	340	7	ADc13605	Adc13605 Human NOV
11	32	91.4	405	3	AAg13173	Aag13173 Arabidops
12	32	91.4	410	8	ADN73063	Adn73063 Thale cre
13	32	91.4	413	3	AAg13173	Aag13173 Arabidops
14	32	91.4	422	3	AAg13172	Aag13172 Arabidops
15	32	91.4	480	7	ADd30464	Addd30464 Plant yie
16	32	91.4	480	8	ADi44439	Adi44439 Plant trea
17	32	91.4	480	8	ADd16268	Addd16268 Thalecres
18	32	91.4	996	5	ABB91788	Abb91788 Herbicide
19	31	88.6	992	3	AAg38672	Aag38672 Arabidops
20	31	88.6	1009	3	AAg38671	Aag38671 Arabidops
21	31	88.6	1027	3	AAg38670	Aag38670 Arabidops
22	30	85.7	99	4	AAU46854	Aau46854 Propionib
23	30	85.7	99	6	ABm43373	Abm43373 Propionib
24	30	85.7	120	4	AAU45935	Aau45935 Propionib
25	30	85.7	120	6	ABm42454	Abm42454 Propionib

26	30	85.7	176	2	AA662775	AA65775	Borrelia
27	30	85.7	190	4	AA662707	AA65707	B. burgdorferi
28	30	85.7	192	2	AA660889	AA66089	Borrelia
29	30	85.7	192	6	ABU61170	ABU6117	Borrelia
30	30	85.7	206	6	ADA14376	ADA1376	Mouse spe
31	30	85.7	283	8	AD624541	AD62541	Bacterial
32	30	85.7	300	4	AA679948	AA679948	Coryneb
33	30	85.7	300	4	AA679801	AA679801	Coryneb
34	30	85.7	300	4	AA690957	AA690957	C. glutami
35	30	85.7	363	7	ADM26713	ADM26713	Hyperther
36	30	85.7	368	4	AA662718	AA662718	Borrelia
37	30	85.7	370	4	AA662715	AA662715	Borrelia
38	30	85.7	374	4	AA662710	AA662710	Borrelia
39	30	85.7	378	4	AA662712	AA662712	Borrelia
40	30	85.7	392	4	AA662734	AA662734	Borrelia
41	30	85.7	394	4	AA662730	AA662730	Borrelia
42	30	85.7	398	4	AA662728	AA662728	Borrelia
43	30	85.7	401	4	AA662738	AA662738	Borrelia
44	30	85.7	415	8	AA625294	AA625294	Bacterial
45	30	85.7	433	7	ABO22403	ABO22403	Mosquito

## ALIGNMENTS

CC	AA02896	standard; peptide; 7 AA.
AC	AA02896;	
DT	18-SEP-2000	(first entry)
DE	Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:37.	
XX		
KW	zinc finger domain; alpha helix; nucleotide binding; DNA binding;	
KW	polydactyl protein; asymmetric target recognition;	
KW	gene specific transcriptional regulator; gene activator; gene repressor;	
KW	transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;	
XX	transgenic animal; antiviral; anticancer; diagnosis.	
OS	Synthetic.	
PN	MO200023464-A2.	
PD	27-APR-2000.	
PF	14-OCT-1999; 99MO-EP007742.	
PR	16-OCT-1998; 98US-00173941.	
PA	(NOVS ) NOVARTIS AG.	
PA	(NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.	
PA	(SCRI ) SCRIPPS RES INST.	
PI	Barbas CF;	
DR	WPI; 2000-339648/29.	
PT	Novel isolated and purified zinc finger nucleotide-binding proteins with	
PT	specificity for GNN triplet sequences, useful in gene therapy and for	
PT	regulating gene function.	
PS	Disclosure; Fig 1; 48pp; English.	
CC	The invention relates to zinc finger nucleotide-binding proteins which	
CC	comprise 2-12, preferably 2-6, operatively linked motifs selected from	
CC	sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha	
CC	helical regions of zinc finger domains which specifically bind to target	
CC	nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked	
CC	by the peptide linker TGERP (AAB02970). The Cys2-His2 zinc finger motif	
CC	is the most frequently utilised nucleic acid binding motif in eukaryotes	
CC	and constitutes a beta-beta-alpha fold. Nucleic acid recognition is	

```

DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein SO1800.
GN OrderedLocustNames=SO1800;
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxId=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813; DOI=10.1038/nbt749;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eilen J.A., Seshadri R., Ward N.L., Methe B.A.,
RA Claydon R.A., Meyer T., Tsapin A., Scott J., Beaman M.J.,
RA Brinkac L.M., DeLong E.C., DeBoy R.T., Dodson R.J., Durkin A.S.,
RA Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A.,
RA White O., Wolf A.M., Vamathevan J.J., Weidman J.F., Uppstam M.,
RA Lee K., Berry K.J., Lee C., Mueller J., Khoult H.M., Gill J.,
RA Uterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O.,
RA Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL: AE015624; AAN54852.1; -.
DR TIGR: SO1800; -.
DR InterPro: IPR010758; SC_ADH.
DR Pfam: PF07055; scADH; 1.
KW Complete proteome.
SQ SEQUENCE 400 AA; 43912 MW; BE230180DA20868 CRC64;

Query Match
Best Local Similarity 100.0%; Score 34; DB 2; Length 400;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 148 DSGELVR 154

RESULT 3
ID Q70453 PRELIMINARY; PRT; 710 AA.
AC Q70453:
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE AGCP11556.
GN Name=AGCG49309; ORFName=ENGANG00000019412;
OS Anopheles gambiae str. PE8T.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxId=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PE8T;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AAB01008964; EAA12458.1; -.
DR HSP: P33222; IUDH.
DR InterPro: IPR008938; ARM.
DR InterPro: IPR000225; Armadillo.
DR InterPro: IPR000357; HEAT.
DR Pfam: PF00514; Arm; 6.
DR Pfam: PF02985; HEAT; 2.
DR PROSITE: PS50176; ARM_REPEAT; 1.
SQ SEQUENCE 710 AA; 77053 MW; B934CD21FA02439 CRC64;

Query Match
Best Local Similarity 100.0%; Score 34; DB 2; Length 710;
Matches 100.0%; Pred. No. 1.1e+02;

```

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 535 DSGELVR 541

RESULT 4
ID Q8K6U1 PRELIMINARY; PRT; 448 AA.
AC Q8K6U1; Q79XM6;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical phage associated protein SPYm3_1130 (Hypothetical protein
DE SP80734).
GN OrderedLocustNames=SP80734, SPYm3_1130;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxId=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;
RA Bers S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu W.-Y., Smoot J.C., Forcella S.F., Perkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence."
RT Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-1;
RX MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okanashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution."
RT Genome Res. 13:1042-1055(2003).
RL EMBL: AE014158; AAM79737.1; -.
DR EMBL: AP005143; BAC63829.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 448 AA; 53074 MW; 9D1F78970DED7A05 CRC64;

Query Match
Best Local Similarity 97.1%; Score 33; DB 2; Length 448;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 361 DSGELVR 367

RESULT 5
ID Q726S1 PRELIMINARY; PRT; 550 AA.
AC Q726S1:
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Methyl-accepting chemotaxis protein, putative.
GN OrderedLocustNames=DVU3035;
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB
OS 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxId=882;
RN [1]

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds

(without alignments)  
51.825 Million cell updates/sec

Title: US-10-646-919-36

Perfect score: 34

Sequence: 1 DSGELVR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03:\*

1: uniprot\_sprot:\*

2: uniprot\_tramb1:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	258	2 Q8BJP8	Q8BJP8 pseudomonas
2	34	100.0	400	2 Q8EG14	Q8EG14 shevanelia
3	34	100.0	710	2 Q7Q453	Q7Q453 anopheles g
4	33	97.1	448	2 Q8K6U1	Q8K6U1 streptococc
5	33	97.1	550	2 Q726S1	Q726S1 desulfovibr
6	31	91.2	91	2 P721S2	P721S2 pseudomonas
7	31	91.2	123	2 Q914N7	Q914N7 pseudomonas
8	31	91.2	132	2 Q61808	Q61808 pseudomonas
9	31	91.2	132	2 Q76EA7	Q76EA7 pseudomonas
10	31	91.2	132	2 Q884Y6	Q884Y6 pseudomonas
11	31	91.2	168	1 GREB_XANCP	GREB_XANCP xanthomonas
12	31	91.2	216	2 Q868N1	Q868N1 ixodes scap
13	31	91.2	299	2 Q9KRA4	Q9KRA4 vibrio chol
14	31	91.2	300	2 Q22090	Q22090 petunia hyb
15	31	91.2	300	2 Q63J16	Q63J16 burkholderi
16	31	91.2	355	2 Q8H5P5	Q8H5P5 oryza sativ
17	31	91.2	386	2 Q6D1E1	Q6D1E1 erwinia car
18	31	91.2	466	2 Q6MP11	Q6MP11 bdellovibr
19	31	91.2	477	1 MURC_XANAC	MURC_XANAC xanthomonas
20	31	91.2	477	1 MURC_XANCP	MURC_XANCP xanthomonas
21	31	91.2	650	2 Q62J76	Q62J76 burkholderi
22	31	91.2	650	2 Q63SW8	Q63SW8 burkholderi
23	31	91.2	685	2 Q9S8S0	Q9S8S0 arabidopsis
24	31	91.2	1079	2 Q758C5	Q758C5 neurospora
25	31	91.2	1286	2 Q752J0	Q752J0 ashbya gos
26	31	91.2	1290	2 Q48756	Q48756 leuconostoc
27	31	91.2	1952	2 Q9S5N5	Q9S5N5 drosophila
28	31	91.2	4547	2 Q9W343	Q9W343 drosophila
29	30	88.2	232	2 Q9BP45	Q9BP45 rhizobium l
30	30	88.2	235	2 Q14195	Q14195 schizosacch
31	30	88.2	266	2 Q88F53	Q88F53 pseudomonas

32	30	88.2	289	2 Q983X2	Q983X2 rhizobium l
33	30	88.2	290	2 Q89P18	Q89P18 bradyrhizob
34	30	88.2	292	2 Q6LNV2	Q6LNV2 photobacter
35	30	88.2	326	2 Q62141	Q62141 burkholderi
36	30	88.2	326	2 Q63RD9	Q63RD9 burkholderi
37	30	88.2	351	2 Q7NV61	Q7NV61 chromobacte
38	30	88.2	367	2 Q97212	Q97212 sulfoblobus
39	30	88.2	447	2 Q15922	Q15922 leishmania
40	30	88.2	543	2 Q97U16	Q97U16 sulfoblobus
41	30	88.2	642	2 Q7ZER9	Q7ZER9 desulfovibr
42	30	88.2	667	2 Q8KW91	Q8KW91 ruegeria sp
43	30	88.2	692	2 DNK1_SYNY3	DNK1_SYNY3
44	30	88.2	861	2 Q88N01	Q88N01 pseudomonas
45	30	88.2	873	1 MJTS_CHRYVO	MJTS_CHRYVO

## ALIGNMENTS

RESULT 1					
Q8BJP8	PRELIMINARY;	PRT;	258 AA.		
ID Q8BJP8:					
DT 01-JUN-2003 (TRMBLrel. 24, Created)					
DT 01-JUN-2003 (TRMBLrel. 24, Last sequence update)					
DT 01-MAR-2004 (TRMBLrel. 26, Last annotation update)					
DE Transcriptional regulator, IclR family.					
DS OrderedLocNames=PP2601;					
OS Pseudomonas putida (strain KT2440).					
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;					
OC Pseudomonadaceae; Pseudomonas.					
OC NCBI_TaxID=160488;					
RP [1]					
RX MEDLINE=22423060; PubMed=12534463;					
RA Nelson K.E., Weibel C., Paulsen I.T., Dodson R.J., Hilbert H.,					
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,					
RA Binkac L.M., Beaman M.J., Deboy R.T., Daugherty S.C., Kolonay J.F.,					
RA Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,					
RA Hance I., Chris Lee P., Holtzapfel B.K., Scanlan D., Tran K.,					
RA Moarrez A., Uterback T.R., Rizzo M., Lee K., Kosack D., Moestl D.,					
RA Wedler H., Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,					
RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuenmler B.,					
RA Frazer C.M.;					
RT "Complete genome sequence and comparative analysis of the					
RL metabolically versatile Pseudomonas putida KT2440.";					
DR Environ. Microbiol. 4:799-808(2002).					
DR EMBL; AE016783; AAA68209.1; -.					
DR TIGR; PP2601; -.					
DR GO; GO:0003677; F:DNA binding; IEA.					
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.					
DR InterPro; IPR005471; HTH_IcIR.					
DR InterPro; IPR009058; HTH_hlx_DNA_bnd.					
DR Pfam; PF01614; IclR; 1.					
DR SMART; SM00346; HTH_IcIR; 1.					
KW Complete proteome; DNA-binding; Transcription;					
KW Transcription regulation.					
SQ SEQUENCE 258 AA; 27864 MW; 541DFDCDEB1772F CRC64;					
Query Match	100.0%;	Score 34;	DB 2;	Length 258;	
Best Local Similarity	100.0%;	Pred. No. 39;			
Matches	7; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	1 DSGELVR 7				
DB	93 DSGELVR 99				
RESULT 2					
ID Q8EG14	PRELIMINARY;	PRT;	400 AA.		
AC Q8EG14;					
DT 01-MAR-2003 (TRMBLrel. 23, Created)					

Query Match 91.2%; Score 31; DB 2; Length 299;  
 Best Local Similarity 85.7%; Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
 |:|||||  
 DB 46 DTGELVR 52

## RESULT 3

T52379  
 zinc finger protein ZPT3-3 [imported] - garden petunia

C:Species: Petunia x hybrida (garden petunia)  
 C:Date: 24-Oct-2000 #sequence\_revision 24-Oct-2000 #text\_change 09-Jul-2004  
 C:Accession: T52379

R:Kudo, K.; Sakamoto, A.; Kobayashi, A.; Rybka, Z.; Kanno, Y.; Nakagawa, H.; Mishino, T.  
 Nucleic Acids Res. 26, 608-615, 1998  
 A:Title: Cys2/His2 zinc-finger protein family of petunia: evolution and general mechanisms  
 A:Reference number: Z26061  
 A:Accession: T52379  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-300 <KUB>

A:Cross-references: UNIPROT:Q02090; EMBL:AB006605; PIDN:BAA21927.1  
 A:Experimental source: strain Mitchell diploid  
 C:Superfamily: Arabidopsis thaliana hypothetical protein F12B4.290

Query Match 91.2%; Score 31; DB 2; Length 300;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
 |:|||||  
 DB 154 DSGELVR 160

## RESULT 4

JCS473  
 dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides

C:Species: Leuconostoc mesenteroides  
 C:Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 09-Jul-2004  
 C:Accession: JCS473  
 R:Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsen, P.  
 Gene 182, 23-32, 1996

A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc  
 A:Reference number: JCS473; MUID:97136686; PMID:8982063  
 A:Accession: JCS473  
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-1290 <MON>  
 A:Cross-references: UNIPROT:Q48756; GB:U38181  
 C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto  
 C:Genetics:  
 A:Gene: dxa  
 C:Keywords: glycosyltransferase; hexosyltransferase  
 F:8-870/Domain: catalytic #status Predicted <CAT>  
 F:922-1290/Domain: glucan-binding #status Predicted <GCB>

Query Match 91.2%; Score 31; DB 2; Length 1290;  
 Best Local Similarity 85.7%; Pred. No. 1,4e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
 |:|||||  
 DB 1217 DTGELVR 1223

## RESULT 5

T38909  
 hypothetical protein SPAC56B4.07 - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T38909  
 R:Comor, R.; Churcher, C.M.; Barrrell, B.G.; Rajandream, M.A.; Wood, V.  
 Submitted to the EMBL Data Library, September 1997  
 A:Reference number: Z21813  
 A:Accession: T38909

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-235 <CON>  
 A:Cross-references: UNIPROT:O14195; EMBL:Z99261; PIDN:CAB13398.1; GSPDB:GN00066; SPDB:SF  
 A:Experimental source: strain 972h-; cosmid G56E4  
 C:Genetics:  
 A:Gene: SPDB:SPAC56B4.07  
 A:Map position: 1  
 C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC56B4.07

Query Match 88.2%; Score 30; DB 2; Length 235;  
 Best Local Similarity 71.4%; Pred. No. 36;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
 |:|||||  
 DB 212 DTGELVR 218

## RESULT 6

C90506  
 hypothetical protein ilvB-6 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
 C:Accession: C90506

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-  
 arett, R.A.; Kagan, M.A.; Sersen, C.W.; Van der Oost, J.  
 Submitted to Genbank, April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: C90506

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-543 <KUR>  
 A:Cross-references: UNIPROT:Q97U16; GB:AE006641; NID:G13816647; PIDN:AAK43306.1; GSPDB:G  
 C:Genetics:  
 A:Gene: ilvB-6

Query Match 88.2%; Score 30; DB 2; Length 543;  
 Best Local Similarity 85.7%; Pred. No. 90;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
 |:|||||  
 DB 129 DTGELVR 135

## RESULT 7

S74372  
 dnaK-type molecular chaperone dnaK - Synechocystis sp. (strain PCC 6803)

N:Alternate names: protein sl10058  
 C:Species: Synechocystis sp.

A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
 C:Accession: S74372  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 s.  
 A:Reference number: S74372; MUID:97061201; PMID:8905231  
 A:Accession: S74372

A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-692 <KAN>  
 A:Cross-references: UNIPROT:O55154; EMBL:D64001; GB:AB001339; NID:G1001102; PIDN:BAA1029  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds  
(without alignments)  
46.449 Million cell updates/sec

Title: US-10-646-919-36

Perfect score: 34  
Sequence: 1 DSGELVR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*\*\*  
1: pir1:\*\*\*  
2: pir2:\*\*\*  
3: pir3:\*\*\*  
4: pir4:\*\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	91.2	123	2 F83507	hypothetical prote
2	31	91.2	299	2 B82164	hypothetical prote
3	31	91.2	300	2 T52379	zinc finger protei
4	31	91.2	1290	2 UC5473	dextranucrase (EC
5	30	88.2	235	2 T38909	hypothetical prote
6	30	88.2	543	2 C90506	hypothetical prote
7	30	88.2	692	2 S74372	dnak-type molecula
8	30	88.2	985	2 B86084	hypothetical prote
9	30	88.2	1377	2 B86034	rhba protein in rh
10	30	88.2	1377	2 C65159	rhba protein in rh
11	30	88.2	1394	2 H91236	rhsh core protein
12	30	88.2	1397	2 A85570	rhsc core protein
13	30	88.2	1397	2 C64805	rhsc core protein
14	30	88.2	1399	2 A99720	rhsc core protein
15	30	88.2	1409	2 F91187	rhba core protein
16	30	88.2	1411	2 B65145	rhsh protein in rh
17	29	85.3	144	2 G70023	hypothetical prote
18	29	85.3	157	2 F95313	hypothetical prote
19	29	85.3	173	2 A95864	hypothetical prote
20	29	85.3	178	2 T09584	high mobility grou
21	29	85.3	184	2 AE2330	adenylate kinase
22	29	85.3	304	2 C89908	hypothetical prote
23	29	85.3	312	2 B96512	hypothetical prote
24	29	85.3	420	2 S53324	glycogen synthase
25	29	85.3	420	1 TVRTKB	tau-protein kinase
26	29	85.3	449	2 S41950	utp-glucose glucos
27	29	85.3	453	2 C86176	hypothetical prote
28	29	85.3	467	2 D88710	protein C43G2.1 [l
29	29	85.3	645	2 F86627	lysr family transc

30	29	85.3	654	2 T32623	hypothetical prote
31	29	85.3	746	2 AF0845	hydrogenase matu
32	29	85.3	777	1 RPBP4	DNA primase - sate
33	29	85.3	777	2 C41830	DNA primase - phag
34	29	85.3	806	2 F69899	phage-related pre-
35	29	85.3	874	2 B70945	hypothetical prote
36	29	85.3	906	2 S35342	Goldi-associated p
37	29	85.3	906	2 S35312	coatomer complex b
38	29	85.3	954	2 T23369	hypothetical prote
39	29	85.3	1179	2 H82706	hypothetical prote
40	29	85.3	2541	2 S11661	talain - mouse
41	28	82.4	53	1 F25G1	photosystem II pro
42	28	82.4	101	2 B38189	protein-tyrosine-p
43	28	82.4	112	2 G86756	prophage p12 prote
44	28	82.4	173	2 B84348	hypothetical prote
45	28	82.4	255	2 G87677	conserved hypothet

## ALIGNMENTS

RESULT 1  
F83507  
hypothetical protein PA1093 [imported] - Pseudomonas aeruginosa (strain PA01)  
C/Spectes: Pseudomonas aeruginosa  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C/Accession: F83507  
R/Stover, C.K.; Pham, X.O.; Ewlin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.D.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Linn, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A/Reference number: A82950; MUID:20437337; PMID:10984043  
A/Accession: F83507  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-123 <STO>  
A/Cross-references: UNIPROT:Q914N7; GB:AE004540; GB:AE004091; NID:G9947008; PIDN:AA60447  
A/Experimental source: strain PA01  
C/Genetics:  
A/Gene: PA1093

Query Match 91.2%; Score 31; DB 2; Length 123;  
Best Local Similarity 85.7%; Pred. No. 10;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
Db 91 DSGELVR 97

RESULT 2  
B82164  
hypothetical protein VC1738 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C/Spectes: Vibrio cholerae  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C/Accession: B82164  
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragoi, I.; Sellers, J.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A/Reference number: A82035; MUID:20406833; PMID:10952301  
A/Accession: B82164  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-259 <HEI>  
A/Cross-references: UNIPROT:Q9KRA4; GB:AE004251; GB:AE003852; NID:G9656248; PIDN:AAF9488  
A/Experimental source: serogroup O1, strain N16961, biotype El Tor  
C/Genetics:  
A/Gene: VC1738  
A/Map position: 1  
C/Superfamily: Xylella faecitiosa hypothetical protein XF1835



**THIS PAGE IS BLANK**

```
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43061
; LENGTH: 985
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (278)..(278)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-43061

Query Match      88.2%; Score 30; DB 15; Length 985;
Best Local Similarity 71.4%; Pred. No. 1.3e+03;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELVR 7
Db      452 DNGELVR 458

RESULT 15
US-10-425-115-312358
; Sequence 312358, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
```

```
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312358
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1017)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_47927C.1.pep
US-10-425-115-312358

Query Match      88.2%; Score 30; DB 16; Length 1017;
Best Local Similarity 85.7%; Pred. No. 1.4e+03;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 DSGELVR 7
Db      312 DSGELVR 318
```

Search completed: July 20, 2005, 14:32:01  
Job time : 72.3333 secs

ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_40278C.1.pep  
US-10-437-963-138936

Query Match 91.2%; Score 31; DB 16; Length 341;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
DB 320 DSGELVR 326

RESULT 11  
US-10-282-122A-47843  
Sequence 47843, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Mang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Foreyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT FILING DATE: 2003-02-20  
PRIOR FILING DATE: 2003-02-20  
PRIOR FILING DATE: 2000-03-21  
PRIOR FILING DATE: 2000-03-21  
PRIOR FILING DATE: 2000-05-23  
PRIOR FILING DATE: 2000-05-23  
PRIOR FILING DATE: 2000-05-26  
PRIOR FILING DATE: 2000-05-26  
PRIOR FILING DATE: 2000-09-06  
PRIOR FILING DATE: 2000-09-06  
PRIOR FILING DATE: 2000-09-09  
PRIOR FILING DATE: 2000-09-09  
PRIOR FILING DATE: 2000-10-23  
PRIOR FILING DATE: 2000-10-23  
PRIOR FILING DATE: 2000-10-23  
PRIOR FILING DATE: 2000-11-27  
PRIOR FILING DATE: 2000-11-27  
PRIOR FILING DATE: 2000-12-22  
PRIOR FILING DATE: 2000-12-22  
PRIOR FILING DATE: 2001-02-09  
PRIOR FILING DATE: 2001-02-09  
PRIOR FILING DATE: 2001-02-16  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 47843  
LENGTH: 624  
TYPE: PRT  
ORGANISM: Burkholderia cepacia  
US-10-282-122A-47843

Query Match 91.2%; Score 31; DB 15; Length 624;  
Best Local Similarity 85.7%; Pred. No. 5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
DB 317 DSGELVR 323

RESULT 12  
US-10-369-493-2805  
Sequence 2805, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT FILING DATE: 2003-02-28  
PRIOR FILING DATE: 2003-02-28  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 2805  
LENGTH: 692  
TYPE: PRT  
ORGANISM: Synechocystis sp.  
US-10-369-493-2805

Query Match 88.2%; Score 30; DB 15; Length 692;  
Best Local Similarity 85.7%; Pred. No. 9.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
DB 547 DSGELVR 553

RESULT 13  
US-10-425-115-312359  
Sequence 312359, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 312359  
LENGTH: 976  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(976)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_47928C.1.pep  
US-10-425-115-312359

Query Match 88.2%; Score 30; DB 16; Length 976;  
Best Local Similarity 85.7%; Pred. No. 1.3e+03;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
DB 312 DSGELVR 318

RESULT 14  
US-10-282-122A-43061  
Sequence 43061, Application US/10282122A  
Publication No. US20040029129A1

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 140877
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_42033C.1.pep
US-10-437-963-140877
```

```
Query Match          91.2%; Score 31; DB 16; Length 79;
Best Local Similarity 85.7%; Pred. No. 56;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
         |||||
Db       35 DSGEVR 41
```

## RESULT 7

```
US-09-864-408A-4672
; Sequence 4672, Application US/09864408A
; Publication No. US2004009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinketsu, Richard A.
; TITLE OF INVENTION: No. US2004009474A1el Human Polynucleotides and Polypeptides Encc
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4672
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-4672
```

```
Query Match          91.2%; Score 31; DB 11; Length 122;
Best Local Similarity 85.7%; Pred. No. 89;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
         |||||
Db       92 DNGELVR 98
```

## RESULT 8

```
US-10-425-114-48472
; Sequence 48472, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
```

```
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48472
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3150-008-C5_F11.pep
US-10-425-114-48472
```

```
Query Match          91.2%; Score 31; DB 15; Length 323;
Best Local Similarity 85.7%; Pred. No. 2,5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
         |||||
Db       300 DAGELVR 306
```

## RESULT 9

```
US-10-425-115-319010
; Sequence 319010, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 319010
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_54009C.1.pep
US-10-425-115-319010
```

```
Query Match          91.2%; Score 31; DB 16; Length 329;
Best Local Similarity 85.7%; Pred. No. 2,5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
         |||||
Db       307 DAGELVR 313
```

## RESULT 10

```
US-10-437-963-138936
; Sequence 138936, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 138936
; LENGTH: 341
; TYPE: PRT
```

```
RESULT 2
US-10-425-115-351857
; Sequence 351857, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_84058C.1.pep
US-10-425-115-351857

Query Match          100.0%; Score 34; DB 16; Length 113;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 96 DSGELVR 102

RESULT 3
US-10-425-114-46116
; Sequence 46116, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701179796_FLI.pep
US-10-425-114-46116

Query Match          100.0%; Score 34; DB 15; Length 125;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 108 DSGELVR 114

RESULT 4
US-10-425-115-351854
; Sequence 351854, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_84058C.1.pep
US-10-425-115-351857

Query Match          100.0%; Score 34; DB 16; Length 113;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 96 DSGELVR 102

RESULT 5
US-10-646-919-35
; Sequence 35, Application US/10646919
; Publication No. US20050148075A1
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos F.
; TITLE OF INVENTION: Zinc-finger binding domains for GNN
; FILE REFERENCE: TSRI 645.2C1
; CURRENT APPLICATION NUMBER: US/10/646,919
; PRIOR FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: US 09/494,190
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/EP99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthesized
US-10-646-919-35

Query Match          91.2%; Score 31; DB 18; Length 7;
Best Local Similarity 85.7%; Pred. No. 1.6e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 1 DSGELVR 7

RESULT 6
US-10-437-963-140877
; Sequence 140877, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(214)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_84055C.1.pep
US-10-425-115-351854

Query Match          100.0%; Score 34; DB 16; Length 214;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 97 DSGELVR 103
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:52:31 ; Search time 71.3333 Seconds  
(without alignments)  
38.093 Million cell updates/sec

Title: US-10-646-919-36  
Perfect score: 34  
Sequence: 1 DSGELVR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10D\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10E\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US10F\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubppaa/US11A\_PUBCOMB.pep:\*  
20: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep:\*  
21: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
22: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	7	18	US-10-646-919-36
2	34	100.0	113	16	US-10-425-115-351857
3	34	100.0	125	15	US-10-425-114-46116
4	34	100.0	214	16	US-10-425-115-351854
5	31	91.2	7	18	US-10-646-919-35
6	31	91.2	79	16	US-10-437-963-140877
7	31	91.2	122	11	US-09-864-408A-4672
8	31	91.2	323	15	US-10-425-114-48472
9	31	91.2	329	16	US-10-425-115-319010
10	31	91.2	341	16	US-10-437-963-138936
11	31	91.2	624	15	US-10-282-122A-47843

	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45																																																																																																																																																																
	30	88.2	692	15	US-10-369-493-2805	30	88.2	976	16	US-10-425-115-312359	30	88.2	985	15	US-10-282-122A-43061	30	88.2	1017	16	US-10-425-115-312358	30	88.2	1377	9	US-09-815-242-10384	30	88.2	1377	14	US-10-287-274-467	30	88.2	1377	15	US-10-282-122A-42731	30	88.2	1377	15	US-10-282-122A-43059	30	88.2	1411	15	US-10-282-122A-43060	30	88.2	5183	14	US-10-107-521-1	22	29	85.3	44	9	US-09-095-881-15	22	29	85.3	60	15	US-10-424-559-215210	24	29	85.3	88	16	US-10-425-115-330300	25	29	85.3	163	16	US-10-437-963-114359	26	29	85.3	167	16	US-10-437-963-105651	27	29	85.3	178	17	US-10-733-923-12994	28	29	85.3	190	16	US-10-437-963-140502	29	29	85.3	253	14	US-10-238-075-1268	30	29	85.3	259	16	US-10-437-963-178622	31	29	85.3	304	15	US-10-282-122A-44419	32	29	85.3	308	9	US-09-815-242-5258	33	29	85.3	319	9	US-09-815-242-12312	34	29	85.3	320	16	US-10-425-115-256929	35	29	85.3	345	16	US-10-437-963-114562	36	29	85.3	350	16	US-10-746-545-27	37	29	85.3	352	16	US-10-664-421-31	38	29	85.3	352	16	US-10-746-545-16	39	29	85.3	352	16	US-10-746-545-17	40	29	85.3	361	9	US-09-916-109-3	41	29	85.3	361	14	US-10-211-412A-3	42	29	85.3	361	16	US-10-450-422-1	43	29	85.3	361	17	US-10-689-461-3	44	29	85.3	367	16	US-10-746-545-32	45	29	85.3	385	16	US-10-664-421-53

#### ALIGNMENTS

RESULT 1  
US-10-646-919-36  
; Sequence 36, Application US/10646919  
; Publication No. US20050148075A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbas, Carlos F.  
; TITLE OF INVENTION: Zinc-finger binding domains for GNN  
; FILE REFERENCE: TSRI 645.2C1  
; CURRENT APPLICATION NUMBER: US/10/646,919  
; PRIOR FILING DATE: 2003-08-21  
; PRIOR APPLICATION NUMBER: US 09/494,190  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: PCT/EP99/07742  
; PRIOR FILING DATE: 1999-10-14  
; PRIOR APPLICATION NUMBER: US 09/173,941  
; PRIOR FILING DATE: 1998-10-16  
; NUMBER OF SEQ ID NOS: 129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 36  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthesized  
US-10-646-919-36

Query Match 100.0%; Score 34; DB 18; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
DB 1 DSGELVR 7

```
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
; OTHER INFORMATION: selected and mutagenized
US-09-494-190-36
```

```
Query Match      100.0%; Score 34; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db      1 DSGDLVR 7
```

```
RESULT 3
US-09-173-941-35
; Sequence 35, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
US-09-173-941-35
```

```
Query Match      91.2%; Score 31; DB 3; Length 7;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db      1 DSGDLVR 7
```

```
RESULT 4
US-09-494-190-35
; Sequence 35, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
US-09-494-190-35
```

```
Query Match      91.2%; Score 31; DB 4; Length 7;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db      1 DSGDLVR 7
```

```
RESULT 5
US-09-711-164-467
; Sequence 467, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Foreyth, R. Allyn
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 1377
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-467
```

```
Query Match      88.2%; Score 30; DB 4; Length 1377;
Best Local Similarity 71.4%; Pred. No. 4.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db      861 DNGELIR 867
```

```
RESULT 6
US-09-095-881-15
; Sequence 15, Application US/09095881
; Patent No. 6489137
; GENERAL INFORMATION:
; APPLICANT: Seeley, Todd
; TITLE OF INVENTION: DETECTION OF LOSS OF THE WILD-TYPE
; FILE REFERENCE: 1405.003 / 200130.437
; CURRENT APPLICATION NUMBER: US/09/095,881
; CURRENT FILING DATE: 1998-06-11
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-095-881-15
```

```
Query Match      85.3%; Score 29; DB 4; Length 44;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELV 6
        |||||
Db      24 DSGELV 29
```

```
RESULT 7
US-09-489-039A-8264
; Sequence 8264, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds  
(without alignments)  
24.117 Million cell updates/sec

Title: US-10-646-919-36

Perfect score: 34

Sequence: 1 DSGELVR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfillseq.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	100.0	7	3	US-09-173-941-36	Sequence 36, Appl
2	34	100.0	7	4	US-09-494-190-36	Sequence 36, Appl
3	31	91.2	7	3	US-09-173-941-35	Sequence 35, Appl
4	31	91.2	7	4	US-09-494-190-35	Sequence 35, Appl
5	30	88.2	1377	4	US-09-711-164-467	Sequence 467, Appl
6	29	85.3	44	4	US-09-095-881-15	Sequence 15, Appl
7	29	85.3	304	4	US-09-489-039A-8264	Sequence 8264, Ap
8	29	85.3	354	4	US-09-540-236-2689	Sequence 2689, Ap
9	29	85.3	361	4	US-09-916-109-3	Sequence 3, Appl1
10	29	85.3	361	4	US-10-211-412B-3	Sequence 3, Appl1
11	29	85.3	394	4	US-09-916-109-2	Sequence 2, Appl1
12	29	85.3	394	4	US-10-211-412B-2	Sequence 2, Appl1
13	29	85.3	420	2	US-08-602-264A-14	Sequence 14, Appl
14	29	85.3	420	4	US-09-316-038-1	Sequence 1, Appl1
15	29	85.3	420	4	US-09-916-109-1	Sequence 1, Appl1
16	29	85.3	420	4	US-10-211-412B-1	Sequence 1, Appl1
17	29	85.3	420	4	US-09-538-092-1163	Sequence 1163, Ap
18	29	85.3	425	4	US-09-270-767-45473	Sequence 45473, A
19	29	85.3	897	4	US-09-802-540-11551	Sequence 11551, A
20	29	85.3	905	4	US-09-538-092-1079	Sequence 1079, Ap
21	29	85.3	906	1	US-08-190-802A-31	Sequence 31, Appl
22	29	85.3	906	1	US-08-477-346-31	Sequence 31, Appl
23	29	85.3	906	3	US-08-473-089-31	Sequence 31, Appl
24	29	85.3	906	4	US-08-487-072A-31	Sequence 31, Appl
25	28	82.4	7	3	US-09-173-941-8	Sequence 8, Appl1
26	28	82.4	7	3	US-09-173-941-37	Sequence 37, Appl
27	28	82.4	7	3	US-09-173-941-77	Sequence 77, Appl

28	28	82.4	7	3	US-09-173-941-107	Sequence 107, App
29	28	82.4	7	4	US-09-494-190-8	Sequence 8, Appl1
30	28	82.4	7	4	US-09-494-190-37	Sequence 37, Appl
31	28	82.4	7	4	US-09-494-190-77	Sequence 77, Appl
32	28	82.4	7	4	US-09-494-190-107	Sequence 107, App
33	28	82.4	101	4	US-09-134-000C-6643	Sequence 6643, Ap
34	28	82.4	115	3	US-09-303-120B-4	Sequence 4, Appl1
35	28	82.4	115	3	US-09-820-576-4	Sequence 4, Appl1
36	28	82.4	115	4	US-09-966-608-4	Sequence 4, Appl1
37	28	82.4	171	4	US-09-270-767-34902	Sequence 34902, A
38	28	82.4	171	4	US-09-270-767-50119	Sequence 50119, A
39	28	82.4	197	4	US-09-107-532A-5430	Sequence 5430, Ap
40	28	82.4	202	3	US-09-171-461-46	Sequence 46, Appl
41	28	82.4	202	4	US-09-970-711-46	Sequence 46, Appl
42	28	82.4	256	4	US-09-252-991A-21603	Sequence 21603, A
43	28	82.4	281	1	US-08-487-748A-9	Sequence 9, Appl1
44	28	82.4	281	1	US-08-487-748A-10	Sequence 10, Appl
45	28	82.4	281	3	US-08-480-070C-10	Sequence 10, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-173-941-36
; Sequence 36, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV0081S
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-36

Query Match      100.0%; Score 34; DB 3; Length 7;
Best local Similarity 100.0%; Pred. No. 4; 1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELVR 7
Db      1 DSGELVR 7

RESULT 2
US-09-494-190-36
; Sequence 36, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
```



ID ADJ98426 standard; peptide; 7 AA.  
 XX  
 AC ADJ98426;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Zinc finger DNA binding peptide #97.  
 XX  
 KM 1library; multimeric DNA binding polypeptide;  
 KM zinc finger DNA binding peptide; gene expression silencing;  
 KM gene expression enhancement.  
 XX  
 OS Unidentified.  
 XX  
 PN MO2003066828-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 PF 07-FEB-2003; 2003MO-US003705.  
 XX  
 PR 07-FEB-2002; 2002US-0354981P.  
 XX  
 PA (SCR1 ) SCRIPPS RES INST.  
 XX  
 PI Barbas CF, Blancafort P;  
 XX  
 DR WPI; 2003-731499/69.  
 XX  
 PT New zinc finger library of multimeric DNA binding polypeptides, useful  
 PT for sterically occluding the binding site of a natural transcription  
 PT factor, and enhancing or silencing target gene expression.  
 XX  
 PS Disclosure; Fig 20; 64pp; English.  
 XX  
 CC The invention comprises a library of multimeric DNA binding polypeptides  
 CC (e.g. zinc finger DNA binding polypeptides). The zinc finger binding  
 CC polypeptides of the invention are useful for sterically occluding the  
 CC binding site of a natural transcription factor and enhancing or silencing  
 CC target gene expression. The present amino acid sequence represents a zinc  
 CC finger DNA binding peptide of the invention.  
 CC  
 SQ Sequence 7 AA;  
 XX  
 QY Query Match 91.2%; Score 31; DB 7; Length 7;  
 Db Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 1 DSGELVR 7  
 1 DSGDLVR 7  
 RESULT 7  
 ABP33363  
 ID ABP33363 standard; protein; 122 AA.  
 XX  
 AC ABP33363;  
 XX  
 DT 08-JUL-2002 (first entry)  
 XX  
 DE Human ORF2336 protein, SEQ ID NO:4672.  
 XX  
 KM Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 KM disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KM immune modulation; haematopoiesis regulation; tissue growth;  
 KM angiogenesis; actinin; inhibin; chemotactic; chemokine; haemostatic;  
 KM thrombolytic; tumour inhibition; bodily characteristics; fertility;  
 KM behaviour; cancer; proliferative disorder; neurological disorder;  
 KM cardiovascular disease; immune system disorder; organ transplantation;  
 KM tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KM hypothyroidism; cholesterol ester storage disease; infection; vulnerability;  
 KM vasoconstrictive; antidiabetic; cytoskeletal; neurotropic;  
 KM neuroprotective; antithrombotic; antidiabetic; antithrombotic; thrombolytic;

KM cardiant; hypotensive; antihypertoid; antiinflammatory; immunomodulator;  
 KM dermatological; analgesic; virucide; antibacterial; fungicide.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200190366-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 24-MAY-2001; 2001MO-US017076.  
 XX  
 PR 24-MAY-2000; 2000US-0206690P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Leach MD, Shinkets RA;  
 XX  
 DR WPI; 2002-106200/14.  
 XX  
 PR N-PSDB; ABN77389.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation.  
 XX  
 PS Claim 10; Page 1421; 2508pp; English.  
 XX  
 CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, actinin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases  
 CC  
 SQ Sequence 122 AA;  
 XX  
 QY Query Match 91.2%; Score 31; DB 5; Length 122;  
 Db Best Local Similarity 85.7%; Pred. No. 65;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 1 DSGELVR 7  
 1 DSGDLVR 98

RESULT 4  
 AAB02894  
 ID AAB02894 standard; peptide; 7 AA.  
 XX  
 AC AAB02894;  
 XX  
 DT 18-SEP-2000 (first entry)  
 XX  
 DE Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:35.  
 XX  
 KM zinc finger domain; alpha helix; nucleotide binding; DNA binding;  
 KM polydactyl protein; asymmetric target recognition;  
 KM gene specific transcriptional regulator; gene activator; gene repressor;  
 KM transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;  
 KM transgenic animal; antiviral; anticancer; diagnosis.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200023464-A2.  
 XX  
 PD 27-APR-2000.  
 XX  
 PF 14-OCT-1999; 99WO-EP007742.  
 XX  
 PR 16-OCT-1998; 98US-00173941.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 PA (SCRI ) SCRIPPS RES INST.  
 XX  
 PI Barbas CF;  
 XX  
 DR MPI; 2000-339648/29.  
 XX  
 PT Novel isolated and purified zinc finger nucleotide-binding proteins with  
 PT specificity for GNN triplet sequences, useful in gene therapy and for  
 PT regulating gene function.  
 XX  
 PS Disclosure; Fig 1; 48pp; English.

The invention relates to zinc finger nucleotide-binding proteins which comprise 2-12, preferably 2-6, operatively linked motifs selected from sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha helical regions of zinc finger domains which specifically bind to target nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked by the peptide linker TGRKP (AAB02970). The Cys2-His2 zinc finger motif is the most frequently utilised nucleic acid binding motif in eukaryotes, and constitutes a beta-beta-alpha fold. Nucleic acid recognition is achieved through specific contacts from side chains of amino acid residues in the alpha helix. Each zinc finger can recognise a sub-site of 3 bp in target DNA. Covalent linkage of multiple zinc finger domains allows the recognition of extended contiguous asymmetric DNA sequences. For example, a synthetic polydactyl protein containing six zinc finger domains can recognise an 18 bp sequence, and such proteins are potentially highly gene-specific. The novel nucleotide-binding zinc finger proteins may therefore be used in the development of artificial gene-specific transcriptional regulators. Such transcriptional switches may be used to regulate the expression of oncogenes such as erbB-2, overexpression of which is involved in malignant transformation. The proteins are therefore useful in the treatment of cancers, and may also be used to activate genes involved in fighting diseases, and to treat viral infections by inhibiting the synthesis of viral gene products. They may be used in DNA-based diagnostic applications. The proteins may also be used in producing functional gene knockout or activation in heterozygous transgenic animals. Proteins of the invention can discriminate between sequences which have a single base difference. This is manifested in a >100-fold decrease in affinity for the variant sequence. Gene activation and repression can be achieved by targeting within the gene transcript, suggesting that information obtained from expressed sequence tags may be sufficient for the construction of gene switches. Sequences AAB02876-B02869 represent zinc finger alpha helix

CC phase library peptides disclosed in the invention  
 XX  
 SQ Sequence 7 AA;  
 XX  
 Query Match 91.2%; Score 31; DB 3; Length 7;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSGELVR 7  
 |||:||||  
 DB 1 DSGDLVR 7  
 |||:||||  
 1 DSGDLVR 7

RESULT 5  
 ADF61984  
 ID ADF61984 standard; peptide; 7 AA.  
 XX  
 AC ADF61984;  
 XX  
 DT 12-FEB-2004 (first entry)  
 XX  
 DE Zinc finger binding region #35.  
 XX  
 KM zinc finger-nucleotide binding polypeptide; expression regulation;  
 KM zinc finger binding region.  
 XX  
 OS Synthetic.  
 XX  
 PN US6610512-B1.  
 XX  
 PD 26-AUG-2003.  
 XX  
 PF 28-JAN-2000; 2000US-00494190.  
 XX  
 PR 16-OCT-1998; 98US-00173941.  
 PR 14-OCT-1999; 99WO-EP007742.  
 XX  
 PA (SCRI ) SCRIPPS RES INST.  
 XX  
 PI Barbas CF;  
 XX  
 DR MPI; 2003-800134/75.  
 XX  
 PT Regulating expression of nucleotide sequence that contains sequence 5'-  
 PT (GNN)n-3', comprising exposing nucleotide sequence to composition  
 PT comprising isolated and purified zinc finger-nucleotide binding  
 PT polypeptide.  
 XX  
 PS Disclosure; SEQ ID NO 35; 46pp; English.

The invention relates to an expression of nucleotide sequence that contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by exposing nucleotide sequence to composition comprising isolated and purified zinc finger-nucleotide binding polypeptide that contains a nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-Arg. The invention is used in the regulation of the expression of the CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present CC sequence represents the amino acid sequence of a zinc finger binding CC region.  
 CC  
 SQ Sequence 7 AA;  
 XX  
 Query Match 91.2%; Score 31; DB 7; Length 7;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSGELVR 7  
 |||:||||  
 DB 1 DSGDLVR 7  
 |||:||||  
 1 DSGDLVR 7

RESULT 6  
 ADJ98426

CC achieved through specific contacts from side chains of amino acid  
CC residues in the alpha helix. Each zinc finger can recognise a subside of  
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains  
CC allows the recognition of extended contiguous asymmetric DNA sequences.  
CC For example, a synthetic polypeptide protein containing six zinc finger  
CC domains can recognise an 18 bp sequence, and such proteins are  
CC potentially highly gene-specific. The novel nucleotide-binding zinc  
CC finger proteins may therefore be used in the development of artificial  
CC gene-specific transcriptional regulators. Such transcriptional switches  
CC may be used to regulate the expression of oncogenes such as *erbB-2*,  
CC overexpression of which is involved in malignant transformation. The  
CC proteins are therefore useful in the treatment of cancers, and may also  
CC be used to activate genes involved in fighting diseases, and to treat  
CC viral infections by inhibiting the synthesis of viral gene products. They  
CC may be used in DNA-based diagnostic applications. The proteins may also  
CC be used in producing functional gene knockout or activation in  
CC heterozygous transgenic animals. Proteins of the invention can  
CC discriminate between sequences which have a single base difference. This  
CC is manifested in a >100-fold decrease in affinity for the variant  
CC sequence. Gene activation and repression can be achieved by targeting  
CC within the gene transcript, suggesting that information obtained from  
CC expressed sequence tags may be sufficient for the construction of gene  
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix  
CC phage library peptides disclosed in the invention  
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 34; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
| | | | |  
Db 1 DSGELVR 7

RESULT 2

ADF61985 standard; peptide; 7 AA.

AC ADF61985;

DT 12-FEB-2004 (first entry)

XX zinc finger binding region #36.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

XX zinc finger binding region.

XX Synthetic.

XX US6610512-B1.

XX 26-AUG-2003.

XX 28-JUN-2000; 2000US-00494190.

XX 16-OCT-1998; 98US-00173941.

XX 14-OCT-1999; 99WO-EP007742.

XX (SCRI ) SCRIPPS RES INST.

XX Barbas CF;

XX WPI; 2003-800134/75.

PT Regulating expression of nucleotide sequence that contains sequence 5'-  
PT (GNN)n-3', comprising exposing nucleotide sequence to composition  
PT comprising isolated and purified zinc finger-nucleotide binding  
PT polypeptide.

XX Disclosure; SEQ ID NO 36; 46pp; English.

CC The invention relates to an expression of nucleotide sequence that  
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by  
CC exposing nucleotide sequence to composition comprising isolated and  
CC purified zinc finger-nucleotide binding polypeptide that contains a  
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Ieu-Val-  
CC Arg. The invention is used in the regulation of the expression of the  
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present  
CC sequence represents the amino acid sequence of a zinc finger binding  
CC region.  
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 34; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
| | | | |  
Db 1 DSGELVR 7

RESULT 3  
ADJ98427 standard; peptide; 7 AA.

AC ADJ98427;

DT 06-MAY-2004 (first entry)

XX zinc finger DNA binding peptide #98.

XX library; multimeric DNA binding polypeptide;

XX zinc finger DNA binding peptide; gene expression silencing;

XX gene expression enhancement.

XX Unidentified.

XX WO2003066828-A2.

XX 14-AUG-2003.

XX 07-FEB-2003; 2003WO-US003705.

XX 07-FEB-2002; 2002US-0354981P.

XX (SCRI ) SCRIPPS RES INST.

XX Barbas CF, Blancafort P;

XX WPI; 2003-731499/69.

XX New zinc finger library of multimeric DNA binding polypeptides, useful  
XX for sterically occluding the binding site of a natural transcription  
XX factor, and enhancing or silencing target gene expression.

XX Disclosure; Fig 20; 64pp; English.

CC The invention comprises a library of multimeric DNA binding polypeptides  
CC (e.g. zinc finger DNA binding polypeptides). The zinc finger binding  
CC polypeptides of the invention are useful for sterically occluding the  
CC binding site of a natural transcription factor and enhancing or silencing  
CC target gene expression. The present amino acid sequence represents a zinc  
CC finger DNA binding peptide of the invention.  
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 34; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7  
| | | | |  
Db 1 DSGELVR 7

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:14:30 ; Search time 82.5 Seconds  
(without alignments)  
32.816 Million cell updates/sec

Title: US-10-646-919-36

Perfect score:	34
Sequence:	1 DSGELVR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match 0%
------------------	------------------

Database : A\_GeneSeq\_16Dec04:\*

```
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	100.0	7	3	AAB02895	Aab02895 Nucleoic acid
2	34	100.0	7	7	ADF61985	Adf61985 zinc finger
3	34	100.0	7	7	ADJ98427	Adj98427 zinc finger
4	31	91.2	7	3	AAB02894	Aab02894 Nucleoic acid
5	31	91.2	7	7	ADF61984	Adf61984 zinc finger
6	31	91.2	7	7	ADJ98426	Adj98426 zinc finger
7	31	91.2	7	7	ABP33363	Abp33363 Human ORF
8	31	91.2	6	6	ABU19919	Abu19919 Protein e
9	31	91.2	4	5	ABBS59051	Abbs5051 Drosophila
10	31	91.2	4	5	AA005784	Aa005784 Human pol
11	30	88.2	125	3	ADK36581	Adk36581 Novel
12	30	88.2	132	5	ABG24355	Abg24355 Novel
13	30	88.2	248	4	ABG07108	Abg07108 Novel
14	30	88.2	285	4	ABG07104	Abg07104 Novel
15	30	88.2	297	4	ABG24354	Abg24354 Novel
16	30	88.2	420	4	ABG08889	Abg08889 Novel
17	30	88.2	420	4	ABG07103	Abg07103 Novel
18	30	88.2	421	4	ABG25615	Abg25615 Novel
19	30	88.2	421	4	ABG35072	Abg35072 Novel
20	30	88.2	421	4	ABG25980	Abg25980 Novel
21	30	88.2	428	4	ABG24361	Abg24361 Novel
22	30	88.2	489	4	ABG08890	Abg08890 Novel
23	30	88.2	489	4	ABG24785	Abg24785 Novel
24	30	88.2	638	4	ABG24363	Abg24363 Novel
25	30	88.2	638	4	ABG08897	Abg08897 Novel

## ALIGNMENTS

[illegible]

Query Match 91.4%; Score 32; DB 2; Length 160;  
 Best Local Similarity 85.7%; Pred. No. 88;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 Db 113 DSGDLVR 119

RESULT 8  
 Q6LBC7 PRELIMINARY; PRT; 167 AA.  
 AC 06LBC7;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 OS Oligotropha carboxidovorans (Pseudomonas carboxydovorans).  
 OC Plasmid pHC3.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Oligotropha.  
 OX NCBI\_TaxID=40137;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OM5;  
 RX MEDLINE=95238294; PubMed=7721710;  
 RA Schubel U., Kraut M., Moersdorf G., Meyer O.;  
 RT "Molecular characterization of the gene cluster *coxMSL* encoding the  
 RT "molybdenum-containing carbon monoxide dehydrogenase of *Oligotropha*  
 RT carboxidovorans.";  
 RL J. Bacteriol. 177:2197-2197(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OM5;  
 RX MEDLINE=97464431; PubMed=9324252;  
 RA Santiago B., Meyer O.;  
 RT "Purification and molecular characterization of the H2 uptake  
 RT membrane-bound Nife-hydrogenase from the carboxidotropic bacterium  
 RT *Oligotropha carboxidovorans*.";  
 RL J. Bacteriol. 179:6053-6060(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OM5;  
 RA Santiago B., Schubel U., Egelseer C., Meyer O.;  
 RT "Sequence analysis, characterization and CO-specific transcription of  
 RT the *cox* gene cluster on the megaplasmid pHC3 of *Oligotropha*  
 RT carboxidovorans.";  
 RL Gene 236:1157-1247(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OM5;  
 RX PubMed=1464498; DOI=10.1016/j.gene.2003.08.027;  
 RA Fuhrmann S., Farmer M., Jeffke T., Henne A., Gottschalk G., Meyer O.;  
 RT "Complete nucleotide sequence of the self-transmissible circular  
 RT megaplasmid pHC3 of *Oligotropha carboxidovorans*. Function in the  
 RT chemolithoautotrophic utilization of CO, H2 and CO2.";  
 RL Gene 322:67-75(2003).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OM5;  
 RA Schubel U.;  
 RT Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OM5;  
 RX Fuhrmann S.;  
 RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: X62447; CAG28432.1; -;  
 KW Hypothetical protein; Plasmid.  
 RT SEQUENCE 167 AA; 18757 MW; F6702F2A6A55553 CRC64;  
 Query Match 91.4%; Score 32; DB 2; Length 167;

Best Local Similarity 85.7%; Pred. No. 93;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 Db 10 DSGDLVR 16

RESULT 9  
 Q8R5H0 PRELIMINARY; PRT; 313 AA.  
 AC Q8R5H0;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE SH2 phosphatase 1 (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mistar;  
 RA Heneberg P., Draber P.;  
 RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF468653; AAL77056.1; -;  
 DR HSSP; P29350; 1GWZ.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0004725; P:protein tyrosine phosphatase activity; IEA.  
 DR GO; GO:000470; P:protein amino acid dephosphorylation; IEA.  
 DR InterPro; IPR000387; TYR\_phosphatase.  
 DR InterPro; IPR000242; Ty\_Pp.  
 DR Pfam; PF00102; Y\_phosphatase; 1.  
 DR PRINTS; PR00700; PRYPHPTASE.  
 DR SMART; SM00194; PTPC; 1.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.  
 KW hydrolase.  
 FT NON TER  
 FT NON TER  
 SQ SEQUENCE 313 AA; 35605 MW; B8A1085FE7737C8B CRC64;

Query Match 91.4%; Score 32; DB 2; Length 313;  
 Best Local Similarity 85.7%; Pred. No. 19e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 Db 119 DSGDLVR 125

RESULT 10  
 Q7WM19 PRELIMINARY; PRT; 452 AA.  
 AC Q7WM19;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Signal-transducing histidine kinase (EC 2.7.3.11).  
 GN Ordered locus names=RB2008;  
 OS *Rhodospirillum rubrum*.  
 OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
 OC Planctomycetaceae; Pirellula.  
 OX NCBI\_TaxID=117;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1;  
 RX MEDLINE=22735913; PubMed=12835416; DOI=10.1073/pnas.1431443100;  
 RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
 RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,  
 RA Schlesner H., Amann R., Reinhardt R.;  
 RT "Complete genome sequence of the marine planctomycete *Pirellula* sp.  
 RT strain 1."

OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFP303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,  
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,  
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti.";  
 RL DNA Res. 7:331-338 (2000).  
 DR EMBL; AP003013; BAB53758.1; -;  
 DR InterPro; IPR008940; Premyl\_trans.  
 DR InterPro; IPR001440; TPR.  
 DR SMART; SM00028; TPR; 4.  
 DR PROSITE; PSS0005; TPR; 2.  
 DR PROSITE; PSS0293; TPR\_REGION; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 289 AA; 3086 MW; D8F9912BD4E87111 CRC64;

Query Match 97.1%; Score 34; DB 2; Length 289;  
 Best Local Similarity 85.7%; Pred. No. 63;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 |||||  
 DB 8 DSGDLIR 14

## RESULT 5

O7RG6 PRELIMINARY; PRT; 400 AA.  
 AC O7RG6; 01-MAR-2004 (TrEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN Name=PY04401;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17XNL;  
 RX PubMed=12368865; DOI=10.1038/nature01099;  
 RA Carlton J.M., Anguioni S.V., Suh B.B., Kooij T.W., Pertea M.,  
 Silva J.C., Ermolaeva M.D., Allen J.E., Sengen J.D., Koo H.L.,  
 Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 Shallom J.J., van Aken S.E., Riedmiller S.B., Feldblyum T.V.,  
 Cho J.K., Quackenbush J., Sedegah M., Shoib A., Cummings L.M.,  
 Florens L., Yates F.R., III, Raine J.D., Sinden R.E., Harris M.A.,  
 Cunningham D.A., Preiser P.R., Bergman L.W., Vaideya A.B.,  
 van Lin L.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 Carucci D.J.;  
 RT "Genome sequence and comparative analysis of the model rodent malaria  
 parasite Plasmodium yoelii yoelii.";  
 RL Nature 419:512-519 (2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AABL0100133; EAA16258.1; -;  
 KW Hypothetical protein.  
 FT NON TER 400  
 FT 400  
 SQ SEQUENCE 400 AA; 47117 MW; DD2PB28296B0303A CRC64;

Query Match 94.3%; Score 33; DB 2; Length 400;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 |||||  
 DB 102 DSGDLVR 108

## RESULT 6

O90U6 PRELIMINARY; PRT; 101 AA.  
 ID O90U6  
 AC O90U6; 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)  
 DE Tyrosine phosphatase from CDNA clone SH-PTP1 (Fragment).  
 OS Rattus sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10118;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92141214; PubMed=1736296;  
 RA Plutsky J., Neel B.G., Rosenberg R.D.;  
 RT "Isolation of a src homology 2-containing tyrosine phosphatase.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:1123-1127 (1992).  
 DR HSP; P29350; IGWZ.  
 DR GO; GO:0004725; P:protein tyrosine phosphatase activity; IEA.  
 DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.  
 DR InterPro; IPR000387; P:protein amino acid dephosphorylation.  
 DR InterPro; IPR000242; Tyr\_PP.  
 DR Pfam; PR00102; Y\_Phasphatase; 1.  
 DR PRINTS; PR00700; PRTYPHRTASE.  
 DR PROSITE; PSS0056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PSS0055; TYR\_PHOSPHATASE\_PTP; 1.  
 FT NON TER 101  
 FT 101  
 SQ SEQUENCE 101 AA; 11404 MW; 04FAF62531D5A3 CRC64;

Query Match 91.4%; Score 32; DB 2; Length 101;  
 Best Local Similarity 85.7%; Pred. No. 63;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 |||||  
 DB 42 DNGDLVR 48

## RESULT 7

O98304 PRELIMINARY; PRT; 160 AA.  
 ID O98304  
 AC O98304; 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)  
 DE M18225 protein.  
 GN Ordered locus names=ml-8225;  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFP303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,  
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,  
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti.";  
 RL DNA Res. 7:331-338 (2000).  
 DR EMBL; AP003013; BAB53826.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 160 AA; 16473 MW; 2F352124BDEDA641 CRC64;

OX NCBI\_TaxId=92829;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=306 / ATCC 13902 / XV 101;  
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,  
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Clapina L.P.,  
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Teai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
RL Nature 417:459-463 (2002).  
CC -1- FUNCTION: Cell wall formation.  
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmutamate + L-alanine = ADP +  
CC phosphate + UDP-N-acetylmutamoyl-L-alanine.  
CC -1- PATHWAY: Peptidoglycan biosynthesis.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -1- SIMILARITY: Belongs to the murCDEP family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; AE011708; AM35668.1; -;  
CC DR HSSP; P45066; IG00.  
CC DR HAMAP; MF\_00046; -; 1.  
CC DR InterPro; IPR004101; Mur\_1lgase\_C.  
CC DR InterPro; IPR000713; Mur\_1lgase\_N.  
CC DR InterPro; IPR005758; MurC.  
CC DR Pfam; PF01225; Mur\_1lgase; 1.  
CC DR Pfam; PF02875; Mur\_1lgase; 1.  
CC DR TIGRPFAM; TIGR01082; murC; 1.  
CC DR ATP-binding; Cell division; Cell wall; Complete proteome; Ligase;  
CC KW NP\_BIND 122 128 ATP (potential).  
CC FT SEQUENCE 477 AA; 50592 MW; 53F60B180C71A2F CRC64;  
SQ  
Query Match 100.0%; Score 35; DB 1; Length 477;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 DSGDLVR 7  
Db 7 DSGDLVR 13  
RESULT 3  
MURC\_XANCP STANDARD; PRT; 477 AA.  
ID MURC\_XANCP  
AC 08PC39;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE UDP-N-acetylmutamate--L-alanine ligase (EC 6.3.2.8) (UDP-N-  
DE acetylmutamoyl-L-alanine synthetase).  
GN Name=murC; OrderedlocusNames=XCC0726;  
OS Xanthomonas campestris (pv. campestris).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxId=340;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33913 / NCPPB 528;  
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,  
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,  
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Clapina L.P.,  
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,  
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Truffi D., Teai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.;  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
RL Nature 417:459-463 (2002).  
CC -1- FUNCTION: Cell wall formation.  
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmutamate + L-alanine = ADP +  
CC phosphate + UDP-N-acetylmutamoyl-L-alanine.  
CC -1- PATHWAY: Peptidoglycan biosynthesis.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -1- SIMILARITY: Belongs to the murCDEP family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; AE012171; AM40041.1; -;  
CC DR HSSP; P45066; IG00.  
CC DR HAMAP; MF\_00046; -; 1.  
CC DR InterPro; IPR004101; Mur\_1lgase\_C.  
CC DR InterPro; IPR000713; Mur\_1lgase\_N.  
CC DR InterPro; IPR005758; MurC.  
CC DR Pfam; PF01225; Mur\_1lgase; 1.  
CC DR Pfam; PF02875; Mur\_1lgase; 1.  
CC DR TIGRPFAM; TIGR01082; murC; 1.  
CC DR ATP-binding; Cell division; Cell wall; Complete proteome; Ligase;  
CC KW NP\_BIND 122 128 ATP (potential).  
CC FT SEQUENCE 477 AA; 50524 MW; 787701B6E2310C89 CRC64;  
SQ  
Query Match 100.0%; Score 35; DB 1; Length 477;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 DSGDLVR 7  
Db 7 DSGDLVR 13  
RESULT 4  
Q983X2 PRELIMINARY; PRT; 289 AA.  
ID Q983X2;  
AC Q983X2;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DE M1R8130 protein.  
GN OrderedlocusNames=m1R8130;  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds  
(without alignments)  
51.825 Million cell updates/sec

Title: US-10-646-919-35

Perfect score: 35

Sequence: 1 DSGDLVR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : uniprot\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	100.0	386	2 Q6DIE1	Q6DIE1 erwina car
2	35	100.0	477	1 MURC_XANAC	Q8PPA7 xanthomonas
3	35	100.0	477	1 MURC_XANCP	Q8PCJ9 xanthomonas
4	34	97.1	289	2 Q983X2	Q983X2 rhizobium 1
5	33	94.3	400	2 Q7RGE6	Q7RGE6 plasmodium
6	32	91.4	101	2 Q9QUU6	Q9QUU6 rattus sp.
7	32	91.4	160	2 Q983Q4	Q983Q4 rhizobium 1
8	32	91.4	167	2 Q6LBC7	Q6LBC7 oligotropha
9	32	91.4	313	2 Q8RSH0	Q8RSH0 rattus norv
10	32	91.4	452	2 Q7UW19	Q7UW19 rhodospirillum
11	32	91.4	595	1 PTN6_MOUSE	P29351 mus musculus
12	32	91.4	613	1 PTN6_RAT	P81718 rattus norv
13	32	91.4	656	2 Q8X438	Q8X438 escherichia
14	32	91.4	682	1 RHSE_ECOLI	P24211 escherichia
15	32	91.4	860	2 Q6D8U4	Q6D8U4 erwina car
16	32	91.4	1398	2 Q7AGW1	Q7AGW1 escherichia
17	32	91.4	1398	2 Q8XE15	Q8XE15 escherichia
18	32	91.4	1400	2 Q8X2F8	Q8X2F8 escherichia
19	32	91.4	1409	2 O52661	O52661 escherichia
20	32	91.4	1426	1 RHSD_ECOLI	P16919 escherichia
21	32	91.4	5289	2 Q88Z03	Q88Z03 lactobacillus
22	31	88.6	106	2 Q7DMC0	Q7DMC0 oryza sativ
23	31	88.6	184	2 Q26364	Q26364 methanobact
24	31	88.6	251	2 Q9L4W2	Q9L4W2 streptomyces
25	31	88.6	258	2 Q88JP8	Q88JP8 pseudomonas
26	31	88.6	263	2 Q72BPF	Q72BPF desulfotomob
27	31	88.6	393	2 Q96RJ7	Q96RJ7 homo sapien
28	31	88.6	393	2 Q96RNO	Q96RNO homo sapien
29	31	88.6	394	2 Q85162	Q85162 pseudomonas
30	31	88.6	394	2 Q9HU25	Q9HU25 pseudomonas
31	31	88.6	400	2 Q8EG14	Q8EG14 shewanella

32	31	88.6	455	1 DD15_STRPU	O17438 strongyloce
33	31	88.6	477	1 MURC_XYLEFA	Q9P180 xylella fas
34	31	88.6	477	1 MURC_XYLEFT	Q87490 xylella fas
35	31	88.6	595	1 PTN6_HUMAN	P29350 homo sapien
36	31	88.6	624	2 Q9UK67	Q9UK67 homo sapien
37	31	88.6	675	2 Q9RUC9	Q9RUC9 delnococtus
38	31	88.6	691	2 Q87019	Q87019 cryptococcus
39	31	88.6	710	2 Q7Q453	Q7Q453 anophelis g
40	31	88.6	879	2 Q7NHE2	Q7NHE2 gloeobacter
41	31	88.6	1006	2 P74686	P74686 synchocyst
42	31	88.6	4467	2 Q9J3B9	Q9J3B9 murine hepa
43	31	88.6	4470	2 Q66WN5	Q66WN5 murine hepa
44	31	88.6	7176	1 R1AB_CYMAS	P16342 m replacase
45	31	88.6	7178	2 Q66WN6	Q66WN6 murine hepa

## ALIGNMENTS

RESULT 1					
ID	Q6DIE1	PRELIMINARY;	PRT;	386 AA.	
AC	Q6DIE1				
DT	25-OCT-2004 (TREMblrel. 28, Created)				
DT	25-OCT-2004 (TREMblrel. 28, Last sequence update)				
DT	25-OCT-2004 (TREMblrel. 28, Last annotation update)				
DE	Putative exported protein.				
GN	OrderedlocusNames=ECA3507;				
OS	Erwinia carotovora (subsp. atroseptica) (Pectobacterium atrosepticum).				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Pectobacterium.				
OX	NCBI_TaxId=29471;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SCRI 1043 / ATCC BAA-672;				
RX	PubMed=15263089; DOI=10.1073/pnas.0402424101.				
RA	Beil K.S., Sebalhia M., Pritchard U., Holden M.T.G., Hyman L.J.,				
RA	Holeva M.C., Thomson N.R., Bentley S.D., Churcher L.J.C., Mungall K.,				
RA	Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,				
RA	Fraser A., Hance Z., Hauser H., Jagsels K., Moule S., Norbertczak H.,				
RA	Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,				
RA	Salmond G.P.C., Birch P.R.U., Parkhill J., Toth I.K.,				
RT	"Genome sequence of the enterobacterial phytopathogen Erwinia				
RT	carotovora subsp. atroseptica and characterization of virulence				
RT	factors."				
RL	Proc. Natl. Acad. Sci. U.S.A. 101:11105-11110 (2004).				
DR	EMBL: BX950851; CAG76405.1; -.				
DR	InterPro: IPR011045; N2O_reductase_N.				
KW	Complete proteome.				
SQ	SEQUENCE 386 AA; 41410 MW; B93562E770DDDE61 CRC64;				
Query Match					
		100.0%;	Score 35;	DB 2;	Length 386;
		Best Local Similarity	100.0%;	Pred. No. 53;	
		Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;	
QY					
	1 DSGDLVR 7				
Db	361 DSGDLVR 367				
RESULT 2					
ID	MURC_XANAC	STANDARD;	PRT;	477 AA.	
AC	Q8PPA7;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	05-JUL-2004 (Rel. 44, Last annotation update)				
DE	UDP-N-acetylglucosamine-6-phosphate-4-epimerase (EC 6.3.2.8) (UDP-N-				
DE	acetylglucosamine-6-phosphate-4-epimerase).				
GN	Name=murc; OrderedlocusNames=XAC0780;				
OS	Xanthomonas axonopodis (pv. citri).				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;				
OC	Xanthomonadaceae; Xanthomonas.				



A:Accession: A42031  
 A:Molecule type: mRNA  
 A:Residues: 1-595 <Y11>  
 A:Cross-references: GB:W68902; NID:g193807  
 A:Note: sequence extracted from NCBI backbone (NCBIN:76845, NCBI:P:76846)  
 R:Yi, T.; Cleveland, J.L.; Ihle, J.N.  
 Blood 78, 2222-2228, 1991  
 A:Title: Identification of novel protein tyrosine phosphatases of hematopoietic cells by  
 A:Reference number: A61180; MUID:92032882; PMID:1932742  
 A:Accession: F61180  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 342-451 <YTA>  
 R:Schultz, L.D.; Schweitzer, P.A.; Rajan, T.V.; Yi, T.; Ihle, J.N.; Matthews, R.J.; Thomas  
 Cell 73, 1445-1454, 1993  
 A:Title: Mutations at the murine motheaten locus are within the hematopoietic cell prote  
 A:Reference number: 152816; MUID:93313972; PMID:8324828  
 A:Accession: 165741  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 334-353, 359-382 <SHU1>  
 A:Cross-references: GB:S63803; NID:g388449  
 A:Note: deletion mutation  
 A:Accession: 152816  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 350-358, 'EGQSPNPLPTPTSSSLVVOYHTQ', 359-366 <SHU2>  
 A:Cross-references: GB:S63763; NID:g388447  
 A:Note: insertion mutation  
 A:Accession: 165740  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 34-76, 'VPRPRHPRAGCVTAAGCGRAD', <SHU3>  
 A:Cross-references: GB:S63764; NID:g388450  
 A:Note: frameshift mutation  
 R:Yeung, Y.G.; Berg, K.L.; Pixley, F.J.; Angelotti, R.H.; Stanley, E.R.  
 J. Biol. Chem. 267, 23447-23450, 1992  
 A:Title: Protein tyrosine phosphatase-1C is rapidly phosphorylated in tyrosine in macro  
 A:Reference number: A45143; MUID:93054686; PMID:1385421  
 A:Accession: A45143  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 137-139, 'X', 141-143, 'X', 145-151 <YE2>  
 A:Experimental source: BAC1.2P5 macrophage (NCBI:P:118519)  
 A:Note: sequence extracted from NCBI backbone (NCBI:P:118519)  
 A:Accession: B45143  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 54-56, 'X', 58, 'X', 60-61, 'X', 63-68 <YE2>  
 A:Experimental source: BAC1.2P5 macrophage  
 A:Note: sequence extracted from NCBI backbone (NCBI:P:118518)  
 C:Comment: This protein is found primarily in hematopoietic tissues.  
 C:Gene: me/HCPH; motheaten  
 C:Function:  
 A:Description: catalyzes hydrolysis of peptidyl-phosphotyrosine to release phosphate  
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-phosph  
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase  
 F:4-98/Domain: SH2 homology <SH2A>  
 F:110-211/Domain: SH2 homology <SH2B>  
 F:265-521/Domain: phosphatase catalytic domain #stratus predicted <P>  
 F:270-504/Domain: protein-tyrosine-phosphatase homology <PTP>  
 F:553/Active site: Cys (phosphocysteine intermediate) #stratus predicted  
 F:459/Binding site: substrate phosphate (Arg) #stratus predicted

Query Match 91.4%; Score 32; DB 1; Length 595;  
 Best Local Similarity 85.7%; Pred. No. 63;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 |:|||||  
 Db 401 DNGDLVR 407

RESULT 3  
 G85731  
 Rhs element associated protein Z2257 [imported] - Escherichia coli (strain O157:H7, sube  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C:Accession: G85731  
 R:Perna, N.T.; Plunkett II, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew  
 Hiler, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouzis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: G85731  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-656 <STO>  
 A:Cross-references: UNIPROT:Q8X438; GB:AE005174; NID:g12515236; PIDN:AA656315.1; GSPDB:G  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Gene: Z2257

Query Match 91.4%; Score 32; DB 2; Length 656;  
 Best Local Similarity 85.7%; Pred. No. 71;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 |:|||||  
 Db 126 DNGDLVR 132

RESULT 4  
 C64898  
 rhaB protein - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004  
 C:Accession: C64898; J06026; S16027  
 R:Blattner, F.R.; Plunkett II, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: C64898  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-662 <BLAT>  
 A:Cross-references: UNIPROT:P24211; GB:AE000242; GB:U00096; NID:g1787720; PIDN:AACT4538  
 A:Experimental source: strain K-12, substrain MG1655  
 R:Sadosky, A.B.; Gray, J.A.; Hill, C.W.  
 Nucleic Acids Res. 19, 7177-7183, 1991  
 A:Title: The RhaD-E subfamily of Escherichia coli K-12.  
 A:Reference number: J06025; MUID:92115567; PMID:1766878  
 A:Accession: J06026  
 A:Molecule type: DNA  
 A:Residues: 9-355, 'X', 357-441, 'G', 443-682 <SA2>  
 A:Cross-references: GB:X60998; NID:g42735; PID:g42736  
 A:Note: the authors translated the codon CAG for residue 349 as Gln  
 C:Gene: rhaB  
 A:Map position: 32 min

Query Match 91.4%; Score 32; DB 2; Length 682;  
 Best Local Similarity 85.7%; Pred. No. 74;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 |:|||||  
 Db 147 DNGDLVR 153

RESULT 5  
 H90698  
 RhaD core protein with extension [imported] - Escherichia coli (strain O157:H7, substra  
 C:Species: Escherichia coli

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

## OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds  
(without alignments)  
46.449 Million cell updates/sec

Title: US-10-646-919-35

Perfect score: 35  
Sequence: 1 DSGDLVR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: PIR\_79:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	91.4	101	2 B38189	protein-tyrosine-P
2	32	91.4	595	1 A44390	protein-tyrosine-P
3	32	91.4	656	2 G85731	Rhs element associ
4	32	91.4	682	2 C64898	Rhs protein - Bsc
5	32	91.4	1398	2 H90698	Rhsd core protein
6	32	91.4	1398	2 B85549	hypothetical prote
7	32	91.4	1400	2 E90886	Rhs core protein
8	32	91.4	1426	2 H64780	rhsp protein precu
9	31	88.6	184	2 C69133	DNA-dependent RNA
10	31	88.6	394	2 D83002	multidrug resistanc
11	31	88.6	477	2 E82763	UDP-N-acetylmurama
12	31	88.6	595	1 S20825	protein-tyrosine-P
13	31	88.6	675	2 E75393	hypothetical prote
14	31	88.6	1006	2 S76892	hypothetical prote
15	30	85.7	315	2 T36417	probable formyltra
16	30	85.7	358	2 F87364	HLyd family secret
17	30	85.7	359	2 T35340	probable adenosine
18	30	85.7	361	2 G95410	probable ABC trans
19	30	85.7	426	2 H70390	conserved hypochet
20	30	85.7	457	2 S44269	platelet-derived g
21	30	85.7	459	2 AC2116	hypothetical prote
22	30	85.7	469	2 S4196	rad22 protein - fl
23	30	85.7	510	2 T13882	cytochrome-c oxida
24	30	85.7	559	2 B95014	metallo-beta-lacta
25	30	85.7	560	2 B86661	hypothetical prote
26	30	85.7	561	2 H69077	DNA ligase - Metha
27	30	85.7	610	2 E97887	conserved hypochet
28	30	85.7	719	2 AC0867	2-acetylglucosylphosp
29	30	85.7	719	2 E91090	hypothetical prote

30	30	85.7	719	2 H85935	hypothetical prote
31	30	85.7	719	2 E65066	2-acetylglucosylphosp
32	30	85.7	788	2 S67595	hypothetical prote
33	30	85.7	920	2 C96831	hypothetical prote
34	30	85.7	926	2 G96563	probable coatomer
35	30	85.7	1353	2 T19691	hypothetical prote
36	30	85.7	2658	2 A86216	protein T23G18.2 (
37	29	82.9	96	2 B85781	hypothetical prote
38	29	82.9	96	2 F90932	hypothetical prote
39	29	82.9	96	2 D64931	hypothetical prote
40	29	82.9	159	2 T37098	hypothetical prote
41	29	82.9	204	2 A10620	probable bacteriop
42	29	82.9	228	2 G72631	hypothetical prote
43	29	82.9	241	2 H81278	phosphatidate cycl
44	29	82.9	262	1 C69097	3',5'-cyclic-nucle
45	29	82.9	275	2 A87487	phosphatidate cycl

## ALIGNMENTS

## RESULT 1

B38189

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 6 - rat (fragment)

N/Alternate names: protein-tyrosine-phosphatase SH-FP1, PTPY42

C/Species: Rattus norvegicus (Norway rat)

C/Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 24-Apr-1998

C/Accession: B38189

R/Plutsky, J.; Neel, B.G.; Rosenberg, R.D.

Proc. Natl. Acad. Sci. U.S.A. 89, 1123-1127, 1992

A/Title: Isolation of a src homology 2-containing tyrosine phosphatase.

A/Reference number: A38189; MUID:92141214; PMID:1736296

A/Accession: B38189

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: nucleic acid

A/Residues: 1-101 <PLU>

A/Experimental source: megakaryocyte

A/Note: sequence extracted from NCBI backbone (NCBIP:79621)

C/Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-phosphatase

C/Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase

F.1-101/Domain: protein-tyrosine-phosphatase homology (fragment) <PTP>

F.1-101/Active site: Cys (phosphocysteine intermediate) #status predicted

F.100/Binding site: substrate phosphate (Arg) #status predicted

Query Match Best local similarity 91.4%; Score 32; DB 2; Length 101;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7

DB 42 DNGDLVR 48

## RESULT 2

A44390

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 6 - mouse

N/Alternate names: hematopoietic cell phosphatase (HCP); protein-tyrosine-phosphatase 1C

C/Species: Mus musculus (house mouse)

C/Date: 03-May-1994 #sequence\_revision 19-May-1994 #text\_change 11-Jun-1999

C/Accession: A44390; A42031; F61180; I65741; I52816; I65740; A45143; B45143

C/Mathews, R.J.; Bowne, D.B.; Flores, E.; Thomas, M.L.

Mol. Cell. Biol. 12, 2396-2405, 1992

A/Title: Characterization of hematopoietic intracellular protein tyrosine phosphatases:

and threonine-rich sequences.

A/Reference number: A44390; MUID:92236615; PMID:1373816

A/Accession: A44390

A/Molecule type: mRNA

A/Residues: 1-595 <MAT>

A/Cross-references: GB:M90389; NTD:9200550; PIDN:AAA40007.1; PID:9200551

R.Yi, T.L.; Cleveland, J.L.; Ihle, J.N.

Mol. Cell. Biol. 12, 836-846, 1992

A/Title: Protein tyrosine phosphatase containing SH2 domains: characterization, preferen

A/Reference number: A42031; MUID:92123209; PMID:1732748

```
RESULT 2
US-10-425-115-354248
; Sequence 354248, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: MRT4577_86246C.1.pep
US-10-425-115-354248

Query Match
Best Local Similarity 91.4%; Score 32; DB 16; Length 112;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
Db 73 DNGDLVR 79

RESULT 3
US-10-366-547-97
; Sequence 97, Application US/10366547
; Publication No. US20030215899A1
; GENERAL INFORMATION:
; APPLICANT: Meng, Tzu-Ching
; APPLICANT: Tonks, Nicholas K.
; APPLICANT: Cool, Deborah E.
; TITLE OF INVENTION: REVERSIBLE OXIDATION OF PROTEIN TYROSINE
; FILE REFERENCE: 200125,439
; CURRENT APPLICATION NUMBER: US/10/366,547
; CURRENT FILING DATE: 2003-02-12
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-366-547-97

Query Match
Best Local Similarity 91.4%; Score 32; DB 15; Length 613;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
Db 403 DNGDLVR 409

RESULT 4
US-10-282-122A-56547
; Sequence 56547, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Hasebeek, Robert
; APPLICANT: Ohlsen, Karl
```

```
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56547
; LENGTH: 682
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-56547

Query Match
Best Local Similarity 91.4%; Score 32; DB 15; Length 682;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
Db 147 DNGDLVR 153

RESULT 5
US-09-912-020-340
; Sequence 340, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DIV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:52:31 ; Search time 71.333 Seconds  
(without alignments)  
38.093 Million cell updates/sec

Title: US-10-646-919-35

Perfect score: 35  
Sequence: 1 DSGDLVR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US6_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US6_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	7	US-10-646-919-35	Sequence 35, Appl
2	32	91.4	112	US-10-425-115-354248	Sequence 354248,
3	32	91.4	613	US-10-366-547-97	Sequence 97, Appl
4	32	91.4	682	US-10-282-122A-56547	Sequence 56547, A
5	32	91.4	1426	US-09-912-020-340	Sequence 340, App
6	32	91.4	1426	US-10-282-122A-42617	Sequence 42617, A
7	32	91.4	1426	US-10-771-241-340	Sequence 340, App
8	31	88.6	7	US-10-646-919-36	Sequence 36, Appl
9	31	88.6	113	US-10-425-115-351857	Sequence 351857,
10	31	88.6	125	US-10-425-114-46116	Sequence 46116, A
11	31	88.6	214	US-10-425-115-351854	Sequence 351854,

12	31	88.6	251	16	US-10-203-295-8	Sequence 8, Appl
13	31	88.6	284	15	US-10-425-114-70219	Sequence 70219, A
14	31	88.6	284	15	US-10-425-115-284122	Sequence 284122,
15	31	88.6	308	15	US-10-425-114-51738	Sequence 51738, A
16	31	88.6	322	9	US-09-788-626-14	Sequence 14, Appl
17	31	88.6	393	15	US-10-094-749-1866	Sequence 1866, Ap
18	31	88.6	595	9	US-09-920-021A-1	Sequence 1, Appl
19	31	88.6	595	15	US-10-366-547-87	Sequence 87, Appl
20	31	88.6	595	15	US-10-366-547-91	Sequence 91, Appl
21	31	88.6	597	15	US-10-366-547-89	Sequence 89, Appl
22	31	88.6	597	18	US-10-491-545A-16	Sequence 16, Appl
23	31	88.6	604	16	US-10-425-115-266821	Sequence 266821,
24	31	88.6	621	15	US-10-264-049-3086	Sequence 3086, Ap
25	31	88.6	843	14	US-10-072-036-117	Sequence 117, App
26	31	88.6	853	14	US-10-072-036-119	Sequence 119, App
27	31	88.6	1006	15	US-10-369-493-2877	Sequence 2877, Ap
28	31	88.6	1006	17	US-10-732-923-16953	Sequence 16953, A
29	30	85.7	89	16	US-10-437-963-175531	Sequence 175531,
30	30	85.7	149	15	US-10-424-539-158436	Sequence 158436,
31	30	85.7	173	15	US-10-424-539-173253	Sequence 173253,
32	30	85.7	236	16	US-10-437-963-156656	Sequence 156656,
33	30	85.7	324	15	US-10-424-539-209732	Sequence 209732,
34	30	85.7	354	14	US-10-156-761-8704	Sequence 8704, Ap
35	30	85.7	509	15	US-10-369-493-9857	Sequence 9857, Ap
36	30	85.7	509	15	US-10-424-539-258996	Sequence 258996,
37	30	85.7	509	15	US-10-425-114-57768	Sequence 57768, A
38	30	85.7	516	16	US-10-425-115-256937	Sequence 256937,
39	30	85.7	559	17	US-10-472-928-42	Sequence 42, Appl
40	30	85.7	560	15	US-10-283-122A-72260	Sequence 72260, A
41	30	85.7	560	15	US-10-282-122A-74738	Sequence 74738, A
42	30	85.7	561	15	US-10-369-493-21517	Sequence 21517, A
43	30	85.7	564	15	US-10-424-539-162243	Sequence 162243,
44	30	85.7	566	15	US-10-424-539-166000	Sequence 166000,
45	30	85.7	574	15	US-10-369-493-3870	Sequence 3870, Ap

## ALIGNMENTS

```
RESULT 1
US-10-646-919-35
; Sequence 35, Application US/10646919
; Publication No. US20050148075A1
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos F.
; TITLE OF INVENTION: Zinc-finger binding domains for GNN
; FILE REFERENCE: TSRI 645.2C1
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: US 09/494,190
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/EP99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthesized
US-10-646-919-35
```

Query Match 100.0%; Score 35; DB 18; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.66+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
DB 1 DSGDLVR 7

Sequence 13, Application US/09144925  
Patent No. 5951979  
GENERAL INFORMATION:  
APPLICANT: Tonks, Nicholas  
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02421-4799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/144,925  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/685,992  
FILING DATE: July 25, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: CSH96-032  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 781-861-6240  
TELEFAX: 781-861-9540  
TELEX:  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 263 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-144-925-13

Query Match 88.6%; Score 31; DB 2; Length 263;  
Best Local Similarity 71.4%; Pred. No. 57;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
|:|||||  
DB 141 DNGDLIR 147

RESULT 10  
US-09-252-991A-23319  
Sequence 23319, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 23319  
LENGTH: 495  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23319  
Query Match 88.6%; Score 31; DB 4; Length 495;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
|:|||||  
DB 174 DDGDLVR 180

RESULT 11  
US-08-202-389-6  
Sequence 6, Application US/08202389  
Patent No. 553636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutsky, Jorge  
APPLICANT: Neel, Benjamin G.  
APPLICANT: Rosenberg, Robert D.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BI992-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 595 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-202-389-6

Query Match 88.6%; Score 31; DB 1; Length 595;  
Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
|:|||||  
DB 401 DNGDLIR 407

RESULT 12

TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN  
FILE REFERENCE: TSRI 645.2  
CURRENT APPLICATION NUMBER: US/09/494,190  
CURRENT FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: EP/99/07742  
PRIOR FILING DATE: 1999-10-14  
PRIOR APPLICATION NUMBER: US 09/173,941  
PRIOR FILING DATE: 1998-10-16  
NUMBER OF SEQ ID NOS: 126  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 36  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: phage display  
OTHER INFORMATION: selected and mutagenized  
US-09-494-190-36

Query Match 88.6%; Score 31; DB 4; Length 7;  
Best Local Similarity 85.7%; Pred. No. 4.1e+05;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
Db 1 DSGELVR 7

## RESULT 7

US-08-202-389-18  
Sequence 18, Application US/08202389  
Patent No. 5536636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutsky, Jorge  
APPLICANT: Neel, Benjamin G.  
APPLICANT: Roseberry, Robert D.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BI92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:  
LENGTH: 97 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-202-389-18

Query Match 88.6%; Score 31; DB 1; Length 97;  
Best Local Similarity 71.4%; Pred. No. 19;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
Db 42 DNGDLVR 48

## RESULT 8

US-08-685-992-13  
Sequence 13, Application US/08685992  
Patent No. 5912138  
GENERAL INFORMATION:  
APPLICANT: Tonks, Nicholas  
APPLICANT: Flint, Andrew J.  
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSER: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,992  
FILING DATE: 25-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
FILING DATE:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: CSHL96-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 781-861-6240  
TELEFAX: 781-861-9540  
TELEX:  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 263 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-685-992-13

Query Match 88.6%; Score 31; DB 2; Length 263;  
Best Local Similarity 71.4%; Pred. No. 57;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
Db 141 DNGDLVR 147

## RESULT 9

US-09-144-925-13

FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:phage display  
OTHER INFORMATION: selected and mutagenized  
US-09-494-190-35

Query Match 100.0%; Score 35; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.1e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
|:|||||  
DB 1 DSGDLVR 7

RESULT 3  
US-08-202-389-2  
Sequence 2, Application US/08202389  
Patent No. 5536636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutzky, Jorge  
APPLICANT: Neel, Benjamin G.  
APPLICANT: Rosenberg, Robert D.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202,389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BIH92-05MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 513 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-202-389-2

Query Match 91.4%; Score 32; DB 1; Length 513;  
Best Local Similarity 85.7%; Pred. No. 72;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DSGDLVR 7  
|:|||||  
DB 319 DNGDLVR 325

RESULT 4  
US-09-492-709A-340  
Sequence 340, Application US/09492709A  
Patent No. 6720139  
GENERAL INFORMATION:  
APPLICANT: Zyskind, Judith  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Trawick, John  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Froelich, Jamie M.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
FILE REFERENCE: ELITRA.001A  
CURRENT APPLICATION NUMBER: US/09/492,709A  
CURRENT FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 340  
LENGTH: 1426  
TYPE: PRT  
ORGANISM: E. COLI  
US-09-492-709A-340

Query Match 91.4%; Score 32; DB 4; Length 1426;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
|:|||||  
DB 872 DNGDLVR 878

RESULT 5  
US-09-173-941-36  
Sequence 36, Application US/09173941  
Patent No. 6140081  
GENERAL INFORMATION:  
APPLICANT: BARBAS, Carlos F.  
TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN  
FILE REFERENCE: NOV0081S  
CURRENT APPLICATION NUMBER: US/09/173,941  
CURRENT FILING DATE: 1998-10-16  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 36  
LENGTH: 7  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: nucleotide  
US-09-173-941-36

Query Match 88.6%; Score 31; DB 3; Length 7;  
Best Local Similarity 85.7%; Pred. No. 4.1e+05;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
|:|||||  
DB 1 DSGDLVR 7

RESULT 6  
US-09-494-190-36  
Sequence 36, Application US/09494190  
Patent No. 6610512  
GENERAL INFORMATION:  
APPLICANT: BARBAS, Carlos F.

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds  
(without alignments)  
24.117 Million cell updates/sec

Title: US-10-646-919-35

Perfect score: 35  
Sequence: 1 DSGDLVR 7

Scoring table: BLOSUM62  
Gapco 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCtUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	7	3 US-09-173-941-35	Sequence 35, Appl
2	35	100.0	7	4 US-09-494-190-35	Sequence 35, Appl
3	32	91.4	513	1 US-08-202-389-2	Sequence 2, Appl
4	32	91.4	1426	4 US-09-492-709A-340	Sequence 340, Appl
5	31	88.6	7	3 US-09-173-941-36	Sequence 36, Appl
6	31	88.6	7	4 US-09-494-190-36	Sequence 36, Appl
7	31	88.6	97	1 US-08-202-389-18	Sequence 18, Appl
8	31	88.6	263	2 US-08-685-992-13	Sequence 13, Appl
9	31	88.6	263	2 US-09-144-925-13	Sequence 13, Appl
10	31	88.6	495	4 US-09-252-991A-23319	Sequence 23319, A
11	31	88.6	595	1 US-08-202-389-6	Sequence 6, Appl
12	31	88.6	621	4 US-09-949-016-8360	Sequence 8360, Ap
13	31	88.6	631	1 US-08-202-389-8	Sequence 8, Appl
14	31	88.6	843	4 US-09-417-197-117	Sequence 117, App
15	31	88.6	853	4 US-09-417-197-119	Sequence 119, App
16	30	85.7	523	4 US-09-555-889A-2	Sequence 2, Appl
17	30	85.7	559	4 US-09-583-110-4863	Sequence 4863, Ap
18	30	85.7	612	3 US-09-107-433-2804	Sequence 2804, Ap
19	29	82.9	7	3 US-09-173-941-50	Sequence 50, Appl
20	29	82.9	7	4 US-09-494-190-50	Sequence 50, Appl
21	29	82.9	162	4 US-09-902-540-9741	Sequence 9741, Ap
22	29	82.9	312	4 US-09-252-991A-19374	Sequence 19374, A
23	29	82.9	332	4 US-09-902-540-9978	Sequence 9978, Ap
24	29	82.9	349	4 US-09-252-991A-27933	Sequence 27933, A
25	29	82.9	401	3 US-08-861-774E-88	Sequence 88, Appl
26	29	82.9	401	4 US-09-489-039A-12491	Sequence 12491, A
27	29	82.9	582	2 US-08-403-852D-20	Sequence 20, Appl

28	29	82.9	582	3 US-08-510-646B-21	Sequence 21, Appl
29	29	82.9	582	3 US-09-231-818-20	Sequence 20, Appl
30	29	82.9	582	4 US-09-635-353B-20	Sequence 20, Appl
31	29	82.9	794	4 US-09-134-000C-5518	Sequence 5518, Ap
32	29	82.9	817	4 US-09-489-039A-10407	Sequence 10407, A
33	29	82.9	927	4 US-09-198-452A-472	Sequence 472, App
34	29	82.9	937	4 US-09-438-185A-449	Sequence 449, App
35	29	82.9	964	4 US-09-902-540-14068	Sequence 14068, A
36	29	82.9	1128	4 US-09-252-991A-31032	Sequence 31032, A
37	29	82.9	1180	4 US-09-252-991A-25722	Sequence 25722, A
38	29	82.9	2323	4 US-09-710-279-760	Sequence 760, App
39	29	82.9	2404	3 US-09-134-001C-3464	Sequence 3464, Ap
40	28	80.0	71	4 US-09-902-540-13751	Sequence 13751, A
41	28	80.0	169	4 US-09-489-039A-11124	Sequence 11124, A
42	28	80.0	171	4 US-09-270-767-34902	Sequence 34902, A
43	28	80.0	235	4 US-09-270-767-50119	Sequence 50119, A
44	28	80.0	235	4 US-09-252-991A-29626	Sequence 29626, A
45	28	80.0	244	4 US-09-248-796A-18903	Sequence 18903, A

#### ALIGNMENTS

```
RESULT 1
US-09-173-941-35
; Sequence 35, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV0081S
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-35

Query Match          100.0%; Score 35; DB 3; Length 7;
Best local similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGDLVR 7
      |||||
Db      1 DSGDLVR 7

RESULT 2
US-09-494-190-35
; Sequence 35, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
```



XX PS Disclosure; Fig 1; 48pp; English.

CC The invention relates to zinc finger nucleotide-binding proteins which  
 CC comprise 2-12, preferably 2-6, operatively linked motifs selected from  
 CC sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha  
 CC helical regions of zinc finger domains which specifically bind to target  
 CC nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked  
 CC by the peptide linker TGERP (AAB02970). The Cys2-His2 zinc finger motif  
 CC is the most frequently utilized nucleic acid binding motif in eukaryotes,  
 CC and constitutes a beta-beta-alpha fold. Nucleic acid recognition is  
 CC achieved through specific contacts from side chains of amino acid  
 CC residues in the alpha helix. Each zinc finger can recognise a sub-site of  
 CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains  
 CC allows the recognition of extended contiguous asymmetric DNA sequences.  
 CC For example, a synthetic polypeptide protein containing six zinc finger  
 CC domains can recognise an 18 bp sequence, and such proteins are  
 CC potentially highly gene-specific. The novel nucleotide-binding zinc  
 CC finger proteins may therefore be used in the development of artificial  
 CC gene-specific transcriptional regulators. Such transcriptional switches  
 CC may be used to regulate the expression of oncogenes such as erbB-2,  
 CC overexpression of which is involved in malignant transformation. The  
 CC proteins are therefore useful in the treatment of cancers, and may also  
 CC be used to activate genes involved in fighting diseases, and to treat  
 CC viral infections by inhibiting the synthesis of viral gene products. They  
 CC may be used in DNA-based diagnostic applications. The proteins may also  
 CC be used in producing functional gene knockout or activation in  
 CC heterozygous transgenic animals. Proteins of the invention can  
 CC discriminate between sequences which have a single base difference. This  
 CC is manifested in a >100-fold decrease in affinity for the variant  
 CC sequence. Gene activation and repression can be achieved by targeting  
 CC within the gene transcript, suggesting that information obtained from  
 CC expressed sequence tags may be sufficient for the construction of gene  
 CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix  
 CC phage library peptides disclosed in the invention  
 CC  
 CC SQ Sequence 7 AA;

Query Match 88.6%; Score 31; DB 3; Length 7;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 |||:||||  
 DB 1 DSGELVR 7

# RESULT 15

ADFE1985  
 ID ADF61985 standard; peptide; 7 AA.

AC ADF61985;

DT 12-FEB-2004 (first entry)

DE Zinc finger binding region #36.

KW zinc finger-nucleotide binding polypeptide; expression regulation;  
 zinc finger binding region.

OS Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

PF 28-JAN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.  
 14-OCT-1999; 99WO-EP007742.

PA (SCRI ) SCRIPPS RES INST.  
 XX

PI Barbas CF;

DR WPI, 2003-800134/75.

PT Regulating expression of nucleotide sequence that contains sequence 5'-  
 (GNN)n-3', comprises exposing nucleotide sequence to composition  
 PT comprising isolated and purified zinc finger-nucleotide binding  
 PT polypeptide.

PS Disclosure; SEQ ID NO 36; 46pp; English.

CC The invention relates to an expression of nucleotide sequence that  
 CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by  
 CC exposing nucleotide sequence to composition comprising isolated and  
 CC purified zinc finger-nucleotide binding polypeptide that contains a  
 CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-  
 CC Arg. The invention is used in the regulation of the expression of the  
 CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present  
 CC sequence represents the amino acid sequence of a zinc finger binding  
 CC region.  
 CC  
 CC SQ Sequence 7 AA;

Query Match 88.6%; Score 31; DB 7; Length 7;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
 |||:||||  
 DB 1 DSGELVR 7

Search completed: July 20, 2005, 13:45:15  
 Job time : 83.5 secs

XX  
CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide  
CC sequences derived from *Escherichia coli* which inhibit *E. coli*  
CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent  
CC nucleotide and protein sequences associated with *E. coli* proliferation.  
CC AAA66056 and AAA66057 represent primers used for sequencing *E. coli*  
CC proliferation inhibiting nucleotide inserts in an example from the  
CC present invention. Methods from the present invention can be used to  
CC identify a proliferation-regulated gene in a microorganism, by contacting  
CC a microorganism with a proliferation-regulated gene activity inhibitory  
CC nucleic acid identified in another organism, and determining if  
CC inhibition occurs in the second microorganism. The nucleic acid sequences  
CC identified as being required for bacterial growth and proliferation, can  
CC be used for antisense therapy for killing bacteria  
XX  
SQ Sequence 1426 AA;  
  
Query Match 91.4%; Score 32; DB 3; Length 1426;  
Best Local Similarity 85.7%; Pred. No. 8.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 DSGDLVR 7  
|:|||||  
Db 872 DNGDLVR 878

RESULT 13  
ABU14693  
ID ABU14693 standard; protein; 1426 AA.  
XX  
AC ABU14693;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #220.  
XX  
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS *Escherichia coli*.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zyskind JW,  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR N-PSDB; ACA18563.  
XX  
DR WPI: 2003-029926/02.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 42617; 1766bp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-regulated gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1426 AA;  
  
Query Match 91.4%; Score 32; DB 6; Length 1426;  
Best Local Similarity 85.7%; Pred. No. 8.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 DSGDLVR 7  
|:|||||  
Db 872 DNGDLVR 878

RESULT 14  
AAB02895  
ID AAB02895 standard; peptide; 7 AA.  
XX  
AC AAB02895;  
XX  
DT 18-SEP-2000 (first entry)  
XX  
DE Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:36.  
XX  
KM Zinc finger domain; alpha helix; nucleotide binding; DNA binding;  
KM polydactyl protein; asymmetric target recognition;  
KM gene specific transcriptional regulator; gene activator; gene repressor;  
KM transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;  
KM transgenic animal; antiviral; anticancer; diagnosis.  
XX  
OS Synthetic.  
XX  
PN WO200023464-A2.  
XX  
PD 27-APR-2000.  
XX  
PF 14-OCT-1999; 99WO-EP007742.  
XX  
PR 16-OCT-1998; 98US-00173941.  
XX  
PA (NOVS ) NOVARTIS AG.  
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
PA (SCRI ) SCRIPPS RES INST.  
XX  
PI Barbas CF;  
XX  
DR WPI: 2000-339648/29.  
XX  
PT Novel isolated and purified zinc finger nucleotide-binding proteins with  
PT specificity for GNN triplet sequences, useful in gene therapy and for  
PT regulating gene function.

KM Food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US008631.  
 PF  
 XX 31-MAR-2000; 2000US-00540217.  
 PR  
 XX 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS79984.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 46156; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 SQ Sequence 973 AA;  
 Query Match 91.4%; Score 32; DB 4; Length 973;  
 Best Local Similarity 85.7%; Pred. No. 5.6e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSGDLVR 7  
 DB 866 DNGDLVR 872  
 RESULT 11  
 ADC01365  
 ID ADC01365 standard; protein; 1400 AA.  
 XX  
 AC ADC01365;  
 XX  
 DT 04-DEC-2003 (first entry)  
 DE Enterohemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 1410.  
 XX  
 KM enterohaemorrhagic; anti-bacterial.  
 XX  
 OS Escherichia coli; O157:H7.

XX  
 PN JP2002355074-A.  
 XX  
 PD 10-DEC-2002.  
 XX  
 PF 24-JAN-2002; 2002JP-00015959.  
 XX  
 PR 24-JAN-2001; 2001JP-00112010.  
 XX  
 PA (UYTS-) UNIV TSUKUBA.  
 XX  
 DR WPI; 2003-451640/43.  
 XX  
 PT Enterohemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule  
 PT and a polypeptide and its use, a polypeptide, a vector and a host cell.  
 XX  
 PS Claim 3; SEQ ID NO 1410; 2067pp; Japanese.  
 XX  
 CC The invention relates to a novel enterohaemorrhagic Escherichia coli  
 CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention  
 CC has anti-bacterial activity. The polypeptide can be used in detection  
 CC and/or treatment of O157:H7 infection. The nucleotide sequence of the  
 CC genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present  
 CC sequence represents an E. coli O157:H7-specific polypeptide of the  
 CC invention.  
 CC  
 SQ Sequence 1400 AA;  
 Query Match 91.4%; Score 32; DB 7; Length 1400;  
 Best Local Similarity 85.7%; Pred. No. 8.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSGDLVR 7  
 DB 870 DNGDLVR 876  
 RESULT 12  
 AAB15983  
 ID AAB15983 standard; protein; 1426 AA.  
 XX  
 AC AAB15983;  
 XX  
 DT 05-OCT-2000 (first entry)  
 DE E. coli proliferation associated protein sequence SEQ ID NO:340.  
 XX  
 KM Escherichia coli; E. coli; proliferation; inhibition; screening;  
 KM antimicrobial; bacterial growth; antisense therapy; antibacterial.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO2000044906-A2.  
 PD 03-AUG-2000.  
 XX  
 PF 27-JAN-2000; 2000WO-US002200.  
 XX  
 PR 27-JAN-1999; 99US-0117405P.  
 XX  
 PA (ELITR-) ELITRA PHARM INC.  
 XX  
 PI Zyskind J, Ohlson KL, Trawick J, Forsyth RA, Froelich JM;  
 PI Carr GJ, Yamamoto RT, Xu HH;  
 DR WPI; 2000-514822/46.  
 DR N-PSDB; AAB65988.  
 XX  
 PT Novel polynucleotides and polypeptides associated with microorganism  
 PT proliferation, used to identify inhibitors of bacterial growth and  
 PT proliferation, for use in antisense therapy.  
 XX  
 PS Claim 11; Page 253-256; 316pp; English.

XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;  
 PI Walli D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA32493.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 56547; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation; (7) identifying a gene in an operon required for  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 SQ Sequence 682 AA;  
 XX  
 Query Match 91.4%; Score 32; DB 6; Length 682;  
 Best Local Similarity 85.7%; Pred. No. 3.8e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSGDLYR 7  
 Db 147 DNGDLYR 153  
 XX  
 RESULT 9  
 ABG09197  
 ID ABG09197 standard; protein; 833 AA.  
 XX  
 AC ABG09197;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #9148.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 KW Homo sapiens.  
 OS WO200175067-A2.  
 PN  
 XX  
 PD 11-OCT-2001.  
 PD  
 PF 30-MAR-2001; 2001WO-US008631.  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSSEQ INC.  
 XX  
 PI Dmanac RT, Liu C, Tang YT;  
 PI  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAG73384.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsiveness for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 39556; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (II) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG0377 represent novel human diagnostic  
 CC amino acid sequences. The present sequence is encoded by one of the  
 CC patent did not appear in the invention. Note: The sequence data for this  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 SQ Sequence 833 AA;  
 XX  
 Query Match 91.4%; Score 32; DB 4; Length 833;  
 Best Local Similarity 85.7%; Pred. No. 4.7e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSGDLYR 7  
 Db 749 DNGDLYR 755  
 XX  
 RESULT 10  
 ABG15797  
 ID ABG15797 standard; protein; 973 AA.  
 XX  
 AC ABG15797;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #15788.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

```

OS   Rattus sp.
XX   US536636-A.
XX   PD
XX   16-JUL-1996.
XX   PF
XX   26-FEB-1994; 94US-00202389.
XX   PR
XX   26-JUN-1991; 91US-00721112.
XX   PR
XX   31-JAN-1992; 92US-00829141.
XX   PR
XX   01-DEC-1992; 92US-00983926.
XX   PA
XX   (BETH-) BETH ISRAEL HOSPITAL ASSOC.
XX   (MAST ) MASSACHUSETTS INST TECHNOLOGY.
XX   PI
XX   Neel BG, Rosenberg RD, Freeman RM, Plutsky J;
XX   WPI; 1996-341506/34.
XX   DR
XX   N-PSDB; AAT35308.
XX   PT
XX   Detecting 12p chromosomal abnormality associated with neoplastic disease
XX   - using SH-PTP1 protein tyrosine phosphatase gene specific probe.
XX   PS
XX   Example 1; Col 32-34; 63pp; English.
XX   CC
XX   AAR9314 is a rat MPTP (protein tyrosine phosphatase-1) clone. A cDNA
XX   clone encoding this sequence was used as a probe to isolate a series of
XX   overlapping human cDNA clones defining the human MPTP equivalent, SH-
XX   PTP1, which contains two tandem SH2 domains. A sequence complementary to
XX   nucleotides 537-653 of the SH-PTP1 gene clone given as AAT35306 is used
XX   as a probe to detect a chromosome 12p3 abnormality associated with
XX   CC
XX   CC neoplastic disease. In partic. acute lymphoblastic leukaemia. The probe
XX   CC hybridises to a part of the region coding for the two tandem SH2 domains
XX   CC (see AAR9312). If the probe will not hybridise DNA of chromosome 12p13
XX   CC from a patient sample it is indicative of an abnormality, normally
XX   CC associated with neoplasia. Fragments of sequences encoding human SH-PTP2
XX   CC (see AAT35307) may also be used to diagnose a condition or susceptibility
XX   CC to a condition associated with chromosome 12p abnormalities.
XX   CC Alternatively the wild-type SH-PTP1 or SH-PTP2 gene or protein may be
XX   CC used for comparison to sequenced PTP genes taken from a patient, where
XX   CC differences indicate an abnormality. The activity of SH-PTP1 may also be
XX   CC purposefully altered by mutation to effect a change in megakaryocyte
XX   CC function and hence platelet production. (Updated on 25-MAR-2003 to
XX   CC correct PF field.)
XX   CC
XX   SQ
XX   Sequence 513 AA:
XX
XX   Query Match . 91.4%; Score 32; DB 2; Length 513;
XX   Best Local Similarity 85.7%; Pred. No. 2.7e+02;
XX   Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX   QY
XX   1 DSGDLVR 7
XX   |:|||||
XX   Db 319 DNGDLVR 325
XX
XX   RESULT 7
XX   ID ADL16248 standard; protein; 613 AA.
XX   AC
XX   ADL16248;
XX   DT
XX   06-MAY-2004 (first entry)
XX   DE
XX   Rat protein tyrosine phosphatase #7.
XX   KM
XX   cytostatic; immunosuppressive; antiallergic;
XX   KM
XX   protein tyrosine phosphatase; reversible oxidation; dephosphorylation;
XX   KM
XX   inducible signalling pathway; cell proliferation; cancer;
XX   KM
XX   guest vs. host disease; autoimmune disease; allergy; metabolic disorder;
XX   cell-cycle abnormality.
XX   OS
XX   Rattus norvegicus.

```

```

XX   MO2003068984-A2.
XX   PN
XX   21-AUG-2003.
XX   PD
XX   13-FEB-2003; 2003WO-EP001446.
XX   PF
XX   13-FEB-2002; 2002US-0356810P.
XX   PR
XX   12-FEB-2003; 2003US-00366547.
XX   PR
XX   (COLD-) COLD SPRING HARBOR LAB.
XX   PA
XX   (CEPT-) CEPTVR INC.
XX   PI
XX   Tonks NK, Tzu-Ching M, Cool DE;
XX   WPI; 2003-712572/67.
XX   DR
XX   N-PSDB; ADL16247.
XX   DR
XX   PT
XX   Identifying reversibly oxidized protein tyrosine phosphatase, useful in
XX   PT screening for specific modulators, potential agents for treating e.g.
XX   PT cancer or autoimmune disease.
XX   PS
XX   Disclosure; SEQ ID NO 97; 238pp; English.
XX   CC
XX   The invention relates to a method for identifying a protein tyrosine
XX   CC phosphatase (PTP) that is reversibly oxidized in a cell by: (i)
XX   CC subducting a sample, including a cell that contains at least one PTP, to
XX   CC conditions that cause reversible oxidation of PTP; (ii) isolating PTP
XX   CC anaerobically, in presence of a sulphydryl-reactive agent (ii) that
XX   CC irreversibly modifies the thiol group of an invariant Cys in the active
XX   CC site of PTP; and (iii) determining, under reducing conditions, the level
XX   CC of dephosphorylation, caused by PTP, of a labelled substrate (iii), where
XX   CC dephosphorylation indicates that an active PTP is present. No details
XX   CC of tests for these activities are given. The method is used to identify
XX   CC reversibly oxidized PTP, also to identify agents that: (a) reversibly
XX   CC modify such PTP; or (b) alter inducible signalling pathways in which PTP
XX   CC are involved. These agents are potentially useful, in human or veterinary
XX   CC medicine, for treating abnormal cell proliferation or growth (cancer);
XX   CC guest vs. host disease; autoimmune diseases; allergy or other
XX   CC immunosuppressed states; metabolic disorders and cell-cycle
XX   CC abnormalities. This sequence represents one of the PTP enzyme of the
XX   CC invention.
XX   CC
XX   SQ
XX   Sequence 613 AA:
XX
XX   Query Match . 91.4%; Score 32; DB 7; Length 613;
XX   Best Local Similarity 85.7%; Pred. No. 3.3e+02;
XX   Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX   QY
XX   1 DSGDLVR 7
XX   |:|||||
XX   Db 403 DNGDLVR 409
XX
XX   RESULT 8
XX   ID ABU28623 standard; protein; 682 AA.
XX   AC
XX   ABU28623;
XX   DT
XX   19-JUN-2003 (first entry)
XX   DE
XX   Protein encoded by Prokaryotic essential gene #14150.
XX   KM
XX   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX   KM
XX   Escherichia coli.
XX   OS
XX   WO200277183-A2.
XX   PN
XX   03-OCT-2002.
XX   PD
XX   21-MAR-2002; 2002WO-US009107.

```

```
RESULT 4
AAR56994
ID AAR56994 standard; protein; 161 AA.
XX
AC AAR56994;
XX
XX 25-MAR-2003 (revised)
DT 28-FEB-1995 (first entry)
XX
XX MPTP-H3/180.
DE
XX Protein tyrosine phosphatase; PTP; primer; PCR; amplification;
KW polymerase chain reaction; restriction analysis; transcription;
KM multigene; reverse transcriptase.
XX
OS Mus musculus.
XX
XX DE4242638-A1.
PN
XX 23-JUN-1994.
PD
XX
XX 17-DEC-1992; 92DE-04242638.
PF
XX 17-DEC-1992; 92DE-04242638.
PR
XX (BOEH/) BOEHM T.
PA
XX Boehm T;
XX
XX MPI; 1994-235494/29.
DR
XX
XX Analysing transcription patterns of gene and multi-gene families - by
PT incomplete PCR with only one labelled primer and restriction analysis of
PT amplification products, also new protein tyrosine phosphate and DNA
PT encoding it.
XX
XX Example 4; Page 14; 22pp; German.
XX
XX Analysis of transcription patterns of gene or multigene families
CC comprises (1) isolating RNA and subjecting it to reverse transcriptase;
CC (2) amplifying the cDNA using one labelled and one unlabelled primer, or
CC primer sets; (3) stopping amplification before conversion of starting
CC materials is complete; and (4) subjecting the DNA produced to restriction
CC analysis. The primer given in AA057682 is used in the amplification of
CC MPTP-H3/180 (given in AAR56994). (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
XX SQ Sequence 161 AA;
XX
XX Query Match 91.4%; Score 32; DB 2; Length 161;
XX Best Local Similarity 85.7%; Pred. No. 73;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 DSGDLVR 7
XX |:|||||
XX Db 104 DNGDLVR 110
XX
XX RESULT 5
XX ABG09201
XX ID ABG09201 standard; protein; 301 AA.
XX
XX AC ABG09201;
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #9192.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
```

```
OS Homo sapiens.
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US008631.
PF
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Dymnac RT, Liu C, Tang YT;
XX
XX MPI; 2001-639362/73.
DR N-FSDB; AAS73386.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 39560; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 301 AA;
XX
XX Query Match 91.4%; Score 32; DB 4; Length 301;
XX Best Local Similarity 85.7%; Pred. No. 1.5e+02;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 DSGDLVR 7
XX |:|||||
XX Db 95 DNGDLVR 101
XX
XX RESULT 6
XX AAR99314
XX ID AAR99314 standard; protein; 513 AA.
XX
XX AC AAR99314;
XX
XX 25-MAR-2003 (revised)
DT 13-NOV-1996 (first entry)
XX
XX Rat MLPTP (protein tyrosine phosphatase-1) clone.
DE
XX PTP; protein tyrosine phosphatase; SH2; Src homology region 2;
KW chromosome 12p; abnormality; mutation; detection; probe; neoplasia;
KW cancer; leukaemia; diagnosis; megakaryocyte regulation.
XX
```

CC achieved through specific contacts from side chains of amino acid  
CC residues in the alpha helix. Each zinc finger can recognise a subsite of  
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains  
CC allows the recognition of extended contiguous asymmetric DNA sequences.  
CC For example, a synthetic polyclonal protein containing six zinc finger  
CC domains can recognise an 18 bp sequence, and such proteins are  
CC potentially highly gene-specific. The novel nucleotide-binding zinc  
CC finger proteins may therefore be used in the development of artificial  
CC gene-specific transcriptional regulators. Such transcriptional switches  
CC may be used to regulate the expression of oncogenes such as erbB-2,  
CC overexpression of which is involved in malignant transformation. The  
CC proteins are therefore useful in the treatment of cancers, and may also  
CC be used to activate genes involved in fighting diseases, and to treat  
CC viral infections by inhibiting the synthesis of viral gene products. They  
CC may be used in DNA-based diagnostic applications. The proteins may also  
CC be used in producing functional gene knockout or activation in  
CC heterozygous transgenic animals. Proteins of the invention can  
CC discriminate between sequences which have a single base difference. This  
CC is manifested in a >100-fold decrease in affinity for the variant  
CC sequence. Gene activation and repression can be achieved by targeting  
CC within the gene transcript, suggesting that information obtained from  
CC expressed sequence tags may be sufficient for the construction of gene  
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix  
CC phage library peptides disclosed in the invention  
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
| | | | |  
Db 1 DSGDLVR 7

RESULT 2  
ADP61984  
ID ADP61984 standard; peptide; 7 AA.

AC ADP61984;

DT 12-FEB-2004 (first entry)

XX zinc finger binding region #35.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

KW zinc finger binding region.

OS Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

PF 28-JAN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.

PR 14-OCT-1999; 99WO-EP007742.

PA (SCRI ) SCRIPPS RES INST.

PI Barbas CF;

DR WPI; 2003-800134/75.

XX Regulating expression of nucleotide sequence that contains sequence 5'-  
PT (GNN)n-3', comprising exposing nucleotide sequence to composition  
PT comprising isolated and purified zinc finger-nucleotide binding  
PT polypeptide.  
XX  
XX Disclosure; SEQ ID NO 35; 46pp; English.  
XX

CC The invention relates to an expression of nucleotide sequence that  
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by  
CC exposing nucleotide sequence to composition comprising isolated and  
CC purified zinc finger-nucleotide binding polypeptide that contains a  
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-  
CC Arg. The invention is used in the regulation of the expression of the  
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present  
CC sequence represents the amino acid sequence of a zinc finger binding  
CC region.  
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
| | | | |  
Db 1 DSGDLVR 7

RESULT 3  
ADJ98426  
ID ADJ98426 standard; peptide; 7 AA.

AC ADJ98426;

DT 06-MAY-2004 (first entry)

DE zinc finger DNA binding peptide #97.

XX library; multimeric DNA binding polypeptide;

KW zinc finger DNA binding peptide; gene expression silencing;

KW gene expression enhancement.

OS Unidentified.

PN WO2003066828-A2.

PD 14-AUG-2003.

PF 07-FEB-2003; 2003WO-US003705.

PR 07-FEB-2002; 2002US-0354981P.

PA (SCRI ) SCRIPPS RES INST.

PI Barbas CF, Blancafort P;

DR WPI; 2003-731499/69.

XX New zinc finger library of multimeric DNA binding polypeptides, useful  
PT for sterically occluding the binding site of a natural transcription  
PT factor, and enhancing or silencing target gene expression.  
XX

PS Disclosure; Fig 20; 64pp; English.

XX The invention comprises a library of multimeric DNA binding polypeptides  
CC (e.g. zinc finger DNA binding polypeptides). The zinc finger binding  
CC polypeptides of the invention are useful for sterically occluding the  
CC binding site of a natural transcription factor and enhancing or silencing  
CC target gene expression. The present amino acid sequence represents a zinc  
CC finger DNA binding peptide of the invention.  
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7  
| | | | |  
Db 1 DSGDLVR 7

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 20, 2005, 13:14:30 ; Search time 82.5 Seconds  
(without alignments)  
32.816 Million cell updates/sec

Title: US-10-646-919-35

Perfect score: 35  
Sequence: 1 DSGDLVR 7Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: A\_Geneseq\_16Dec04:\*  
2: geneseqp1980a:\*  
3: geneseqp1990a:\*  
4: geneseqp2000a:\*  
5: geneseqp2001a:\*  
6: geneseqp2002a:\*  
7: geneseqp2003a:\*  
8: geneseqp2004a:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	7	3 AAB02894	Aab02894 Nucleotid
2	35	100.0	7	7 ADF61984	Adf61984 Zinc fing
3	35	100.0	7	7 ADJ98426	Adj98426 Zinc fing
4	32	91.4	161	2 AAB56994	Aab56994 WPTP-H3/1
5	32	91.4	301	4 ABOG9201	Abog9201 Novel hum
6	32	91.4	513	2 AAB99314	Aab99314 Rat MLPTP
7	32	91.4	613	7 ADL16248	Adl16248 Rat prote
8	32	91.4	682	6 ABU28623	Abu28623 Protein e
9	32	91.4	833	4 ABOG9197	Abog9197 Novel hum
10	32	91.4	973	4 ABOG15797	Abog15797 Novel hum
11	32	91.4	1400	7 ADC01365	Adc01365 Enterocae
12	32	91.4	1426	3 AAB15983	Aab15983 E. coli p
13	32	91.4	1426	6 ABU14693	Abu14693 Protein e
14	31	88.6	7	3 AAB02895	Aab02895 Nucleotid
15	31	88.6	7	7 ADF61985	Adf61985 Zinc fing
16	31	88.6	7	7 ADJ98427	Adj98427 Zinc fing
17	31	88.6	234	3 AAB51882	Aab51882 Gene 5 hu
18	31	88.6	251	4 AAE10130	Aae10130 Streptomy
19	31	88.6	263	4 AAB59376	Aab59376 Human pro
20	31	88.6	299	3 AAB34205	Aab34205 Gene 41 h
21	31	88.6	322	4 AAG78274	Aag78274 Human STL
22	31	88.6	365	5 ADM19697	Adm19697 Protein e
23	31	88.6	393	5 AAB57882	Aab57882 Human RNA
24	31	88.6	393	6 ADA54298	Ada54298 Human pro
25	31	88.6	495	7 ABO74573	Ab074573 Pseudomon

26	31	88.6	518	4 AAB59241	Aab59241 SHP-1 pro
27	31	88.6	556	7 ADG74664	Adg74664 Human kin
28	31	88.6	556	8 AABM83634	Aabm83634 Human dia
29	31	88.6	595	2 AAB59312	Aab59312 Human SH-
30	31	88.6	595	4 AAB59236	Aab59236 SHP-1 act
31	31	88.6	595	4 AAB59235	Aab59235 SHP-1 act
32	31	88.6	595	4 AAB59228	Aab59228 SHP-1 act
33	31	88.6	595	4 AAB59233	Aab59233 SHP-1 act
34	31	88.6	595	4 AAB59216	Aab59216 SHP-1 pro
35	31	88.6	595	4 AAB59239	Aab59239 SHP-1 E74
36	31	88.6	595	4 AAB59237	Aab59237 SHP-1 act
37	31	88.6	595	4 AAB59231	Aab59231 SHP-1 act
38	31	88.6	595	4 AAB59232	Aab59232 SHP-1 act
39	31	88.6	595	4 AAB59230	Aab59230 SHP-1 act
40	31	88.6	595	4 AAB59238	Aab59238 SHP-1 D59
41	31	88.6	595	4 AAB59229	Aab59229 SHP-1 act
42	31	88.6	595	7 ADL16238	Adl16238 Human pro
43	31	88.6	595	7 ADL16242	Adl16242 Human pro
44	31	88.6	595	8 ADH50796	Adh50796 Human SHP
45	31	88.6	595	8 ADH50796	Adh50796 Human SHP

## ALIGNMENTS

RESULT 1	
AAB02894	
ID	AAB02894 standard; peptide; 7 AA.
XX	
AC	AAB02894;
XX	
DT	18-SEP-2000 (first entry)
XX	
DE	Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:35.
XX	
KW	Zinc finger domain; alpha helix; nucleotide binding; DNA binding;
KW	poly(acyl) protein; asymmetric target recognition;
KW	gene specific transcriptional regulator; gene activator; gene repressor;
KW	transcriptional switch; oncogene; erbA-2; cancer; tumour; gene therapy;
KW	transgenic animal; antiviral; anticancer; diagnosis.
XX	
OS	Synthetic.
XX	
PN	WO200023464-A2.
XX	
PD	27-APR-2000.
XX	
PF	14-OCT-1999; 99WO-EP007742.
XX	
PR	16-OCT-1998; 98US-00173941.
XX	
PA	(NOVS ) NOVARTIS AG.
PA	(NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PA	(SCRI ) SCRIPPS RES INST.
XX	
PI	Barbas CF;
XX	
DR	WPI; 2000-339648/29.
XX	
PT	Novel isolated and purified zinc finger nucleotide-binding proteins with
PT	specificity for GNN triplet sequences, useful in gene therapy and for
PT	regulating gene function.
XX	
PS	Disclosure; Fig 1; 48pp; English.
XX	
CC	The invention relates to zinc finger nucleotide-binding proteins which
CC	comprise 2-12, preferably 2-6, operatively linked motifs selected from
CC	sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha
CC	helical regions of zinc finger domains which specifically bind to target
CC	nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked
CC	by the peptide linker TGRK (AAB02970). The Cys2-His2 zinc finger motif
CC	is the most frequently utilised nucleic acid binding motif in eukaryotes,
CC	and constitutes a beta-beta-alpha fold. Nucleic acid recognition is



RA Ameterdam A., Nissen R.M., Sun Z., Swindell E.C., Farrington S.,  
 RA Hopkins N.;  
 RT "identification of 315 genes essential for early zebrafish  
 RT development.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:12792-12797(2004).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=whole;  
 RX MEDLINE=22388557; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Streuberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stagleon M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughlano N.A., Peters G.J., Abraham R.D., Mullaly S.J.,  
 RA Boeak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gamarate P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Hulton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywniak M.I., Skalska U., Smaltz D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=whole;  
 RX Director MGC Project;  
 RL Submitted (JUL-2004) to the EMBL/Genbank/DBSJ databases.  
 DR EMBL; AY648824; AAT68142.1; -  
 DR EMBL; BC076020; AAH76020.1; -  
 DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.  
 DR InterPro; IPR004854; UFD1.  
 DR Pfam; PF03152; UFD1; 1.  
 SQ SEQUENCE 308 AA; 35000 MW; F5A890BPAAJ1ADBA CRC64;

Query Match 94.6%; Score 35; DB 2; Length 308;  
 Best Local Similarity 85.7%; Pred. No. 59;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DPSPDKR 7  
 |||||  
 Db 251 DPSPDKR 257

RESULT 3  
 06ZIG6  
 ID 06ZIG6 PRELIMINARY; PRT; 763 AA.  
 AC 06ZIG6;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Putative ZfpA.  
 GN Name=OJ1115\_B01.25; Synonyms=OJ1073\_F05.14;  
 OS Oryza sativa (Japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Euphorbiaceae; Oryzae; Oryza.  
 OC NCB1\_Taxid=39947;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RA Sasaki T., Matsunoto T., Yamamoto K.;  
 RA Submitted (AUG-2001) to the EMBL/Genbank/DBSJ databases.  
 DR EMBL; AP004000; BAD15444.1; -  
 DR EMBL; AP003990; BAD15414.1; -  
 SQ SEQUENCE 763 AA; 86146 MW; B5787A807577F693 CRC64;

94.6%; Score 35; DB 2; Length 763;  
 Query Match

	Best Local Similarity	85.7%	Pred. No.	1.6e+02;	Mismatches	0;	Gaps	0;
	Matches	6;	Conservative	1;	Mismatches	0;	Indels	0;
Oy	1 DPSDLK R      :							
Db	653 DPSDKR	659						
RESULT 4								
ID	Q9RZAS	PRELIMINARY;	PRT;	149 AA.				
AC	Q9RZA5.							
DT	01-MAY-2000 (TREMBLrel. 13,	Created)						
DJ	01-MAY-2000 (TREMBLrel. 13,	Last sequence update)						
DE	Response regulator.							
CN	OrderedLocustNames=DDA0049;							
OS	Deinococcus radiodurans.							
OC	Bacteria; Deinococcus-Thermus; Deinococcici; Deinocecales;							
NB	NCBI_TaxID=1299;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;							
RX	MEDLINE=20036896; PubMed=10567266; DOI=10.1126/science.286.5444.1571;							
RA	White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,							
RA	Dodonson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,							
RA	Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,							
RA	Vamathevan J.J., Lam P., McDonald L.A., Utterback T.R., Zalewski C.,							
RA	Makarov A.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,							
RA	Ketchum K.A., Nelson K.E., Salzberg S.L., Smith H.O., Venter J.C.,							
RA	Frazer C.M.;							
RT	"Genome sequence of the radioresistant bacterium Deinococcus							
RL	radiodurans R1."							
DR	Science 286:1571-1577(1999).							
DR	EMB1; AE001862; AAFI262.I; -.							
DR	PIR; C75598; C75598.							
DR	HSSP; Q5169; I13C.							
DR	TIGR; DRAD0049; "-							
DR	GO; GO:0003677; F:DNA binding; IEA.							
DR	GO; GO:0000156; F:two-component response regulator activity; IEA.							
DR	GO; GO:0007600; F:sensory perception; IEA.							
DR	GO; GO:0000160; P:two-component signal transduction system (p...); IEA.							
DR	InferPro; IPRO011006; CheY like.							
DR	InterPro; IPRO011789; Response_reg.							
DR	Pfam; PF000072; Response_reg; 1.							
DR	Prodom; PD000039; Response_reg; 1.							
DR	SMART; SMART0448; REC; 1.							
DR	PROSITE; PS00110; RESPONSE_REGULATORY; 1.							
SQ	Complete proteome; Phosphorylation; Sensory transduction.							
KW	Sequence_149 AA; 16604 MW; 6530AEBA957246A CRC64;							
Query Match								
Oy	1 DPSDLK R      :	91.9%;	Score 34;	DB 2;	Length 149;			
Db	100 DPSPVKR	106	Best local similarity 85.7%; Pred. NO. 42;					
			Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;					
RESULT 5								
ID	Q9RXO0	PRELIMINARY;	PRT;	166 AA.				
AC	Q9RXO0.							
DT	01-MAY-2000 (TREMBLrel. 13,	Created)						
DJ	01-MAY-2000 (TREMBLrel. 13,	Last sequence update)						
DE	01-MAR-2004 (TREMBLrel. 26,	Last annotation update)						
CN	Hypothetical protein DR0259.							
OS	OrderedListNames=DR0259;							
OC	Deinococcus radiodurans.							
OC	Bacteria; Deinococcus-Thermus; Deinococcici; Deinocecales;							
OC	Deinococcaceae; Deinococcus.							

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 seconds

(without alignments)  
51.825 Million cell updates/sec

Title: US-10-646-919-34

Perfect score: 37

Sequence: 1 DPSDLKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 segs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot\_03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	100.0	727	2	Q9CPJ7 pasteurella
2	35	94.6	308	2	Q6DRD5 brachydanio
3	35	94.6	763	2	Q6ZIG6 oryza sativ
4	34	91.9	149	2	Q9RZAS deinococcus
5	34	91.9	168	2	Q9RXQ0 aeropyrum p
6	34	91.9	277	2	Q9FUD3 arabidopsis
7	34	91.9	392	2	Q9YA73 ashyba goss
8	34	91.9	798	2	Q75AG7 klyveromyc
9	34	91.9	812	2	Q6CL61 arabidopsis
10	34	91.9	985	2	Q04387 saccharomyc
11	34	91.9	1032	2	Q80491 arabidopsis
12	34	91.9	1035	2	Q6N6N9 rhodospirillum rubrum
13	34	91.9	1116	2	Q8CHD2 mus musculus
14	34	91.9	1163	2	Q94872 homo sapien
15	34	91.9	1585	1	I4G3 HUMAN
16	34	91.9	1623	2	Q6RZX2 equus caball
17	33	89.2	111	2	Q9YF60 aeropyrum p
18	33	89.2	153	2	Q9RTU3 delinococcus
19	33	89.2	210	2	Q6R2U7 mus musculus
20	33	89.2	215	2	Q22845 arabidopsis
21	33	89.2	265	2	Q8VC24 mus musculus
22	33	89.2	309	2	Q8WTV4 m cytochrome
23	33	89.2	335	2	Q22595 griffithia s
24	33	89.2	356	2	Q8LYD0 methanopyru
25	33	89.2	412	1	KAPR SCHPO
26	33	89.2	1353	2	Q62UM1 schizosacch
27	33	89.2	1356	2	Q63V28 burkholderi
28	33	86.5	138	2	Q8F926 leptospira
29	32	86.5	186	2	Q8KV86 uncultured
30	32	86.5	268	2	Q978S4 thermoplasma
31	32	86.5	304	2	Q9CU23 mus musculus

32	32	86.5	310	2	Q984W5 rhizobium 1
33	32	86.5	334	1	G3P PYRAB
34	32	86.5	334	1	G3P PYRFU
35	32	86.5	334	1	G3P PYRMO
36	32	86.5	336	1	SYW_GLOVI
37	32	86.5	340	2	Q9E299
38	32	86.5	344	2	Q9A9F8
39	32	86.5	354	1	SYW_AGR75
40	32	86.5	354	1	SYW_RHIME
41	32	86.5	355	1	SYW_BRUME
42	32	86.5	355	1	SYW_BRUSU
43	32	86.5	355	1	SYW_RHTLO
44	32	86.5	358	2	Q8C915
45	32	86.5	381	2	Q99XN6

## ALIGNMENTS

RESULT 1	ID	Q9CPJ7	PRELIMINARY;	PRT;	727 AA.
AC	Q9CPJ7				
DT	01-JUN-2001 (TREMBLrel. 17, Created)				
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)				
DE	pFhr.				
GN	Name=pFhr; OrderedLocusNames=PM0040;				
OS	Pasteurella multocida.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;				
OC	Pasteurellaceae; Pasteurella.				
NCBI_TaxID=747;					
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Em70;				
RA	MEDLINE=21145866; PubMed=11248100; DOI=10.1073/pnas.051634598;				
RT	May B.U., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;				
RT	"Complete genomic sequence of Pasteurella multocida pm70."				
RT	Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).				
DR	EMBL; AE006037; AK02124.1; -				
DR	GO; GO:0019867; C:outer membrane; IEA.				
DR	GO; GO:0048772; F:receptor activity; IEA.				
DR	GO; GO:0005215; F:transporter activity; IEA.				
DR	GO; GO:0006810; P:transport; IEA.				
KW	Complete proteome.				
SO	SEQUENCE 727 AA; 81332 MW; 714A089F8F60ADE CRC64;				
Query Match					
Best Local Similarity	100.0%; Score 37; DB 2; Length 727;				
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 DPSDLKR 7				
DB	126 DPSDLKR 132				
RESULT 2					
Q6DRD5	PRELIMINARY;	PRT;	308 AA.		
AC	Q6DRD5;				
DT	25-OCT-2004 (TREMBLrel. 28, Created)				
DT	25-OCT-2004 (TREMBLrel. 28, Last sequence update)				
DT	25-OCT-2004 (TREMBLrel. 28, Last annotation update)				
DE	Ubiquitin fusion degradation 1-like.				
GN	Name=Ubl1;				
OS	Brachydanio rerio (Zebrafish) (Danio rerio).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;				
OC	Cyprinidae; Danio.				
NCBI_TaxID=7955;					
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	PubMed=15256591; DOI=10.1073/pnas.0403929101;				



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds

(without alignments)  
46.449 Million cell updates/sec

Title: US-10-646-919-34

Perfect score: 37

Sequence: 1 DPSDLKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	91.9	149	2 C75598	response regulator
2	34	91.9	168	2 C75541	conserved hypothe
3	34	91.9	392	1 D72511	probable hexosyltr
4	34	91.9	985	2 S15965	hypothetical prote
5	34	91.9	1032	2 B86224	hypothetical prote
6	33	89.2	111	2 D72730	hypothetical prote
7	33	89.2	153	2 B75367	hypothetical prote
8	33	89.2	169	2 D84864	probable calcium b
9	33	89.2	347	2 T39140	camp-dependent pro
10	33	89.2	411	2 S18634	cgsl protein - fis
11	32	86.5	334	1 DE0YG	glyceralddehyde 3-P
12	32	86.5	344	2 E75153	hypothetical prote
13	32	86.5	344	2 B87377	hypothetical prote
14	32	86.5	354	2 AE2619	tryptophan-tRNA 11
15	32	86.5	383	2 E97401	tryptophan-tRNA 11
16	32	86.5	385	2 AH3477	tryptophan-tRNA 11
17	32	86.5	393	2 S16321	light-induced prot
18	32	86.5	409	2 C87538	conserved hypothe
19	32	86.5	855	2 T41336	probable nitrogen
20	32	86.5	2352	2 T43431	alpha-glucan synth
21	31	83.8	145	2 F70514	hypothetical prote
22	31	83.8	312	2 T26085	hypothetical prote
23	31	83.8	403	2 T47621	bZIP transcription
24	31	83.8	588	2 B83507	probable ATP-bindi
25	31	83.8	694	2 JC7673	dynein intermediat
26	31	83.8	706	2 G90696	hypothetical prote
27	31	83.8	708	2 JC6329	yeast secretory pr
28	31	83.8	720	2 C85547	probable cytoplasm
29	31	83.8	732	2 T31629	hypothetical prote

30	31	83.8	1036	2 E96682	hypothetical prote
31	31	83.8	1158	2 T25082	hypothetical prote
32	30	81.1	150	2 H86194	hypothetical prote
33	30	81.1	153	2 D84706	hypothetical prote
34	30	81.1	155	2 H69029	mutator Muir relat
35	30	81.1	192	2 T21210	hypothetical prote
36	30	81.1	216	1 ZZZRBL	hypothetical prote
37	30	81.1	271	2 T31314	modulation protein
38	30	81.1	277	2 F71335	probable A/G-speci
39	30	81.1	281	2 S33023	probable uracil-DN
40	30	81.1	366	1 S35189	dihydrokaempferol
41	30	81.1	401	2 T41049	tyrosine-tRNA liga
42	30	81.1	440	2 G87444	8-amino-7-oxonon
43	30	81.1	451	2 T35914	probable carboxype
44	30	81.1	455	2 F71153	probable Vi polysa
45	30	81.1	519	2 S66673	disulfide isomeras

#### ALIGNMENTS

RESULT 1  
C75598  
Response regulator - Deinococcus radiodurans (strain R1)  
C/Species: Deinococcus radiodurans  
C/Date: 03-Dec-1999 #sequence\_rev1510 03-Dec-1999 #ext\_change 16-Aug-2004  
C/Accession: C75598  
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Uitterbeck, T.; Zalewski, C.; M  
Science 286, 1571-1577, 1999  
A/Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.  
A/Reference number: A75250; PMID:10567266  
A/Accession: C75598  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-149 <WHI>  
A/Cross-references: UNIPROT:Q9R2A5; GB:AE001862; GB:AE001825; NID:96460468; PIDN:AAF122  
A/Experimental source: strain R1  
C/Genetics:  
A/Gene: DR0049  
A/Map position: 2  
C/Superfamily: Signal transduction receiver (phosphoacceptor) protein, CheY type, respo

Query Match 91.9%; Score 34; DB 2; Length 149;  
Best Local Similarity 85.7%; Pred. No. 8;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7  
Db 100 DPSDLKR 106  
|||||  
|||

RESULT 2  
C75541  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C/Species: Deinococcus radiodurans  
C/Date: 03-Dec-1999 #sequence\_rev1510 03-Dec-1999 #ext\_change 09-Jul-2004  
C/Accession: C75541  
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Uitterbeck, T.; Zalewski, C.; M  
Science 286, 1571-1577, 1999  
A/Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.  
A/Reference number: A75250; PMID:10567266  
A/Accession: C75541  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-166 <WHI>  
A/Cross-references: UNIPROT:Q9R2A5; GB:AE001862; GB:AE000513; NID:96457928; PIDN:AAF098  
A/Experimental source: strain R1  
C/Genetics:  
A/Gene: DR0259

RESULT 2  
US-10-437-963-173601  
; Sequence 173601, Application US/10437963  
; Publication No. US2004012343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 173601  
; LENGTH: 763  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_71622C.1.pep  
US-10-437-963-173601

Query Match 94.6%; Score 35; DB 16; Length 763;  
Best Local Similarity 85.7%; Pred. No. 2.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7  
|:|:|:|  
Db 653 DPSDLR 659

RESULT 3  
US-10-425-115-359614  
; Sequence 359614, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 359614  
; LENGTH: 112  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_91140C.1.pep  
US-10-425-115-359614

Query Match 91.9%; Score 34; DB 16; Length 112;  
Best Local Similarity 85.7%; Pred. No. 57;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7  
|:|:|:|  
Db 48 DPSDLR 54

RESULT 4  
US-10-424-599-220780  
; Sequence 220780, Application US/10424599  
; Publication No. US20040031072A1

; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 220780  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_41395C.1.pep  
US-10-424-599-220780

Query Match 91.9%; Score 34; DB 15; Length 283;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7  
|:|:|:|  
Db 118 DPSDLR 124

RESULT 5  
US-10-369-493-22940  
; Sequence 22940, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22940  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Aeropyrum pernix  
US-10-369-493-22940

Query Match 91.9%; Score 34; DB 15; Length 392;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7  
|:|:|:|  
Db 91 DPSDLR 97

RESULT 6  
US-10-120-801-71  
; Sequence 71, Application US/10120801  
; Publication No. US20030203843A1  
; GENERAL INFORMATION:  
; APPLICANT: Pena, Carol  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Shinkets, Richard  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Mehraban, Fuad



FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:phage display  
; OTHER INFORMATION: selected and mutagenized  
US-09-494-190-34

Query Match 100.0%; Score 37; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4,1e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7  
Db 1 DPSDLR 7

RESULT 3  
US-09-543-681A-6003  
; Sequence 6003, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETTON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709,1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 6003  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-6003

Query Match 91.9%; Score 34; DB 4; Length 364;  
Best Local Similarity 85.7%; Pred. No. 41;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7  
Db 215 DPSDLR 221

RESULT 4  
US-09-248-796A-26549  
; Sequence 26549, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; FILE REFERENCE: 107196,132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 26549  
; LENGTH: 311  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-26549

Query Match 86.5%; Score 32; DB 4; Length 311;  
Best Local Similarity 85.7%; Pred. No. 88;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPSDLR 7  
Db 235 DPSDLR 241

RESULT 5  
US-09-949-016-6838  
; Sequence 6838, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6838  
; LENGTH: 396  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-6838

Query Match 83.8%; Score 31; DB 4; Length 396;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSDLR 7  
Db 315 PSDLR 320

RESULT 6  
US-09-949-016-10646  
; Sequence 10646, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10646  
; LENGTH: 409  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-10646

Query Match 83.8%; Score 31; DB 4; Length 409;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSDLR 7  
Db 328 PSDLR 333

RESULT 7  
US-09-252-991A-22913  
; Sequence 22913, Application US/09252991A  
; Patent No. 6551795

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds  
(without alignments)  
24.117 Million cell updates/sec

Title: US-10-646-919-34

Perfect score: 37

Sequence: 1 DSDDKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	100.0	7	US-09-173-941-34	Sequence 34, Appl
2	37	100.0	7	US-09-494-190-34	Sequence 34, Appl
3	34	91.9	364	US-09-543-681A-6003	Sequence 6003, Ap
4	32	86.5	311	US-09-248-796A-26549	Sequence 26549, A
5	31	83.8	396	US-09-949-016-6838	Sequence 6838, Ap
6	31	83.8	409	US-09-949-016-10646	Sequence 10646, A
7	31	83.8	440	US-09-252-991A-22913	Sequence 22913, A
8	31	83.8	475	US-09-252-991A-31373	Sequence 31373, A
9	31	83.8	508	US-09-252-991A-19543	Sequence 19543, A
10	31	83.8	723	US-09-328-352-6928	Sequence 369, App
11	31	83.8	708	US-09-643-597-369	Sequence 19543, A
12	31	83.8	818	US-09-252-991A-16691	Sequence 16691, A
13	31	83.8	906	US-09-252-991A-32715	Sequence 32715, A
14	31	83.8	1089	US-09-252-991A-20334	Sequence 20334, A
15	30	81.1	7	US-09-173-941-33	Sequence 33, Appl
16	30	81.1	7	US-09-494-190-33	Sequence 33, Appl
17	30	81.1	108	US-09-540-236-3176	Sequence 3176, Ap
18	30	81.1	271	US-09-248-796A-16882	Sequence 16882, A
19	30	81.1	271	US-09-408-020-16	Sequence 16, Appl
20	30	81.1	272	US-09-177-165A-27	Sequence 27, Appl
21	30	81.1	315	US-09-252-991A-19219	Sequence 19219, A
22	30	81.1	366	US-09-638-715-2	Sequence 2, Appl1
23	30	81.1	366	US-09-638-715-4	Sequence 2, Appl1
24	30	81.1	366	US-10-060-509-2	Sequence 2, Appl1
25	30	81.1	366	US-10-060-509-4	Sequence 4, Appl1
26	30	81.1	366	US-10-060-506-2	Sequence 2, Appl1
27	30	81.1	366	US-10-060-506-4	Sequence 4, Appl1

28	30	81.1	497	4	US-09-252-991A-30012	Sequence 30012, A
29	30	81.1	788	4	US-09-252-991A-17380	Sequence 17380, A
30	30	81.1	815	4	US-09-177-165A-24	Sequence 24, Appl
31	30	81.1	912	4	US-09-252-991A-24163	Sequence 24163, A
32	29	78.4	90	4	US-09-270-767-32706	Sequence 32706, A
33	29	78.4	118	4	US-09-248-796A-28195	Sequence 28195, A
34	29	78.4	185	4	US-09-562-737-115	Sequence 115, App
35	29	78.4	209	2	US-08-870-518-10	Sequence 10, Appl
36	29	78.4	328	4	US-09-679-279-9	Sequence 9, Appl
37	29	78.4	370	4	US-09-134-000C-6618	Sequence 6618, Ap
38	29	78.4	378	4	US-09-252-991A-23259	Sequence 23259, A
39	29	78.4	459	1	US-08-630-592-2	Sequence 2, Appl1
40	29	78.4	459	1	US-08-714-991-2	Sequence 2, Appl1
41	29	78.4	459	2	US-08-870-518-1	Sequence 1, Appl1
42	29	78.4	459	3	US-09-032-365A-2	Sequence 2, Appl1
43	29	78.4	460	1	US-08-630-592-7	Sequence 7, Appl1
44	29	78.4	460	1	US-08-714-991-7	Sequence 7, Appl1
45	29	78.4	460	3	US-09-032-365A-8	Sequence 8, Appl1

#### ALIGNMENTS

```

RESULT 1
US-09-173-941-34
; Sequence 34, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; US-09-173-941-34

Query Match      100.0%; Score 37; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSDDKR 7
      |||||
Db      1 DSDDKR 7

RESULT 2
US-09-494-190-34
; Sequence 34, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence

```



KW pathogen tolerance; pest tolerance; plant disease resistance;  
 KW cell cycle pathway modification; plant growth regulator;  
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
 KW bacterial polypeptide.  
 XX  
 OS Bacteria.  
 PN US2003233675-A1.  
 PD 18-DEC-2003.  
 PF 20-FEB-2003; 2003US-00369493.  
 PR 21-FEB-2002; 2002US-0360039P.  
 XX  
 PA (CAOY/) CAO Y.  
 PA (HINK/) HINKLE G J.  
 PA (SLAT/) SLATER S C.  
 PA (CHER/) CHEN X.  
 PA (GOLD/) GOLDMAN B S.  
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 DR WPI; 2004-061375/06.  
 XX  
 PT New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.  
 XX  
 PS Claim 1; SEQ ID NO 22940; 122pp; English.  
 XX  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polypeptide used in the  
 CC scope of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
 XX  
 SQ Sequence 392 AA;  
 XX  
 Query Match 91.9%; Score 34; DB 8; Length 392;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DPSDLKR 7  
 Db 91 DPSDLRR 97  
 XX  
 RESULT 7  
 ADG39839 standard; protein; 901 AA.  
 XX  
 AC ADG39839;

XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Protein similar to human NOV8 #5.  
 XX  
 KW Human; NOVX; cardiomyopathy; atherosclerosis; hypertension;  
 KW congenital heart defect; pulmonary stenosis; scleroderma; obesity;  
 KW metabolic disturbance; obesity; transplantation; adrenoleukodystrophy;  
 KW congenital adrenal hyperplasia; prostate cancer; diabetes;  
 KW metabolic disorder; neoplasm; adenocarcinoma; fertility; haemophilia;  
 KW graft versus host disease; AIDS; bronchial asthma; Crohn's disease;  
 KW multiple sclerosis; infectious disease; anorexia;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW immune disorder; haematopoietic disorder; dyslipidaemia;  
 KW wasting disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003203843-A1.  
 PD 30-OCT-2003.  
 PF 11-APR-2002; 2002US-00120801.  
 XX  
 PR 20-APR-2001; 2001US-0285609P.  
 PR 23-APR-2001; 2001US-0285748P.  
 PR 24-APR-2001; 2001US-0286068P.  
 PR 25-APR-2001; 2001US-0286292P.  
 PR 03-MAY-2001; 2001US-0288334P.  
 PR 16-MAY-2001; 2001US-0291241P.  
 PR 14-SEP-2001; 2001US-0322284P.  
 XX  
 PA (PENA/) PENA C E A.  
 PA (GUOX/) GUO X.  
 PA (SHIM/) SHIMKETS R A.  
 PA (PADI/) PADIGARU M.  
 PA (KEKU/) KEKUDA R.  
 PA (SPYT/) SPYTEK K A.  
 PA (MEHR/) MEHRABAN F.  
 PA (TOPP/) TOPPER J N.  
 PA (MALY/) MALYANKAR U M.  
 PA (WASS/) WASSERMAN S M.  
 PA (EDIN/) EDINGER S R.  
 PA (SMIT/) SMITHSON G.  
 PA (GUNT/) GUNTHER E.  
 PA (KOMU/) KOMUVES L.  
 XX  
 PI Pena CE, Guo X, Shimkets RA, Padigaru M, Kekuda R, Spyteck KA;  
 PI Mehraban F, Topper JN, Malyankar UM, Wasserman SM, Edinger SR;  
 PI Smithson G, Gunther E, Komuves L;  
 DR WPI; 2003-900671/82.  
 XX  
 PT New NOVX polypeptides and nucleic acids, useful for diagnosing or  
 PT treating e.g. cardiomyopathy, atherosclerosis, hypertension, scleroderma,  
 PT obesity, prostate cancer, AIDS, bronchial asthma, Crohn's disease, or  
 PT multiple sclerosis.  
 XX  
 PS Disclosure; SEQ ID NO 71; 215pp; English.  
 XX  
 CC The invention relates to a new isolated polypeptide comprising an amino  
 CC acid sequence selected from 17 fully defined human NOVX sequences (even  
 CC SEQ ID NOS between ADG39770 and ADG39802), a mature form of the NOVX  
 CC amino acid or a variant of NOVX, where one or more amino acid residue in  
 CC the variant differs in no more than 15% of the amino acid residues of  
 CC NOVX. Also included are an isolated nucleic acid (NA) molecule  
 CC (comprising a nucleic acid sequence encoding a NOVX polypeptide above  
 CC (odd SEQ ID NOS between ADG39769 and ADG39801), a nucleic acid fragment  
 CC encoding at least a portion of a NOVX polypeptide and a complement of  
 CC NOVX NA), a vector comprising NOVX NA, a cell comprising the vector, an  
 CC antibody that immunospecifically binds to NOVX, a method for determining  
 CC the presence or amount of NOVX or NOVX NA in a sample, a method of  
 CC identifying an agent that binds to NOVX, a method for identifying an

Db 100 DPSDLKK 106

[illegible]

CC achieved through specific contacts from side chains of amino acid  
CC residues in the alpha helix. Each zinc finger can recognise a subunit of  
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains  
CC allows the recognition of extended contiguous asymmetric DNA sequences.  
CC For example, a synthetic polynucleotide protein containing six zinc finger  
CC domains can recognise an 18 bp sequence, and such proteins are  
CC potentially highly gene-specific. The novel nucleotide-binding zinc  
CC finger proteins may therefore be used in the development of artificial  
CC gene-specific transcriptional regulators. Such transcriptional switches  
CC may be used to regulate the expression of oncogenes such as erbB-2,  
CC overexpression of which is involved in malignant transformation. The  
CC proteins are therefore useful in the treatment of cancers, and may also  
CC be used to activate genes involved in fighting diseases, and to treat  
CC viral infections by inhibiting the synthesis of viral gene products. They  
CC may be used in DNA-based diagnostic applications. The proteins may also  
CC be used in producing functional gene knockout or activation in  
CC heterozygous transgenic animals. Proteins of the invention can  
CC discriminate between sequences which have a single base difference. This  
CC is manifested in a >100-fold decrease in affinity for the variant  
CC sequence. Gene activation and repression can be achieved by targeting  
CC within the gene transcript, suggesting that information obtained from  
CC expressed sequence tags may be sufficient for the construction of gene  
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix  
CC phase library peptides disclosed in the invention  
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 37; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPDDLKR 7  
Db 1 DPDDLKR 7

RESULT 2  
ADP61983  
ID ADF61983 standard; peptide; 7 AA.

AC ADF61983;

DT 12-FEB-2004 (first entry)

XX Zinc finger binding region #34.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

XX zinc finger binding region.

OS Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

XX 28-JAN-2000; 2000US-00494190.

XX 16-OCT-1998; 98US-00173941.

PR 14-OCT-1999; 99WO-EP007742.

PA (SCRI ) SCRIPPS RES INST.

XX Barbas CF;

PI Barbas CF;

XX MPI; 2003-800134/75.

XX Regulating expression of nucleotide sequence that contains sequence 5'-  
XX (GNN)n-3', comprises exposing nucleotide sequence to composition  
XX comprising isolated and purified zinc finger-nucleotide binding  
XX polypeptide.  
XX disclosure; SEQ ID NO 34; 46pp; English.

CC The invention relates to an expression of nucleotide sequence that  
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by  
CC exposing nucleotide sequence to composition comprising isolated and  
CC purified zinc finger-nucleotide binding polypeptide that contains a  
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Ileu-Val-  
CC Arg. The invention is used in the regulation of the expression of the  
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present  
CC sequence represents the amino acid sequence of a zinc finger binding  
CC region.  
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 37; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPDDLKR 7  
Db 1 DPDDLKR 7

RESULT 3  
ADJ98425  
ID ADJ98425 standard; peptide; 7 AA.

AC ADJ98425;

DT 06-MAY-2004 (first entry)

DE Zinc finger DNA binding peptide #96.

XX library; multimeric DNA binding polypeptide;

XX zinc finger DNA binding peptide; gene expression silencing;

XX gene expression enhancement.

XX Unidentified.

XX WO200306828-A2.

XX 14-AUG-2003.

XX 07-FEB-2003; 2003WO-US003705.

XX 07-FEB-2002; 2002US-0354981P.

XX (SCRI ) SCRIPPS RES INST.

XX Barbas CF, Blancafort P;

XX MPI; 2003-731499/69.

XX New zinc finger library of multimeric DNA binding polypeptides, useful  
XX for sterically occluding the binding site of a natural transcription  
XX factor, and enhancing or silencing target gene expression.  
XX disclosure; Fig 20; 64pp; English.

XX The invention comprises a library of multimeric DNA binding polypeptides  
XX (e.g. zinc finger DNA binding polypeptides). The zinc finger binding  
XX polypeptides of the invention are useful for sterically occluding the  
XX binding site of a natural transcription factor and enhancing or silencing  
XX target gene expression. The present amino acid sequence represents a zinc  
XX finger DNA binding peptide of the invention.  
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 37; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPDDLKR 7  
Db 1 DPDDLKR 7

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:14:30 ; Search time 82.5 Seconds  
(without alignments)  
32.816 Million cell updates/sec

Title: US-10-646-919-34

Perfect score: 37  
Sequence: 1 DPSDKR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	100.0	7	3 AAB02893	Aab02893 Nucleotid
2	37	100.0	7	7 ADP61983	Adf61983 Zinc fing
3	37	100.0	7	7 ADJ98425	Adj98425 Zinc fing
4	34	91.9	178	4 ABG20437	Abg20437 Novel hum
5	34	91.9	364	7 ADF05718	Adf05718 Bacterial
6	34	91.9	352	8 ADS44510	Ads44510 Bacterial
7	34	91.9	901	7 ADG39839	Adg39839 Protein s
8	34	91.9	1139	8 ADI16228	Adi16228 Human nuc
9	34	91.9	1163	7 ADG39838	Adg39838 Protein s
10	34	91.9	1556	8 ADG97964	Adg97964 Human can
11	34	91.9	1584	3 AAY84430	Hay84430 Amino aci
12	34	91.9	1585	8 ADG97962	Adg97962 Human can
13	34	91.9	1756	4 ABG20440	Abg20440 Novel hum
14	33	89.2	169	3 AAG43188	Aag43188 Arabidops
15	33	89.2	169	3 AAG07957	Aag07957 Arabidops
16	33	89.2	215	3 AAG43187	Aag43187 Arabidops
17	33	89.2	215	3 AAG07956	Aag07956 Arabidops
18	33	89.2	215	8 ADN73189	Adn73189 Thale cire
19	33	89.2	309	7 ADM25766	Adm25766 Mouse inh
20	33	89.2	356	7 ADM25766	Adm25766 Hyperther
21	33	89.2	752	6 ABU1254	Abu1254 Protein e
22	33	89.2	1319	6 ABU22579	Abu22579 Protein e
23	32	86.5	179	6 ABU46358	Abu46358 Protein e
24	32	86.5	314	4 AAB96123	Aab96123 Putative
25	32	86.5	387	5 ABP25701	Abp25701 Streptoco

26	32	86.5	459	8 ADG44904	Adg44904 Mouse bac
27	32	86.5	468	5 AAE24155	Aae24155 Human 466
28	32	86.5	468	5 ADG44908	Adg44908 Human sac
29	32	86.5	571	8 ADP04715	Adp04715 Sea equit
30	32	86.5	676	8 ADG30434	Adg30434 Bacterial
31	31	83.8	158	4 ABG13584	Abg13584 Novel hum
32	31	83.8	205	4 ABG25323	Abg25323 Novel hum
33	31	83.8	298	7 ADM25590	Adm25590 Hyperther
34	31	83.8	357	7 ADJ69162	Adj69162 Human hea
35	31	83.8	396	6 ABH82854	Abh82854 SPO11 rel
36	31	83.8	396	6 ABH82855	Abh82855 SPO11 rel
37	31	83.8	403	8 ADI61353	Adi61353 A. thalia
38	31	83.8	403	8 ADI02359	Adi02359 Thalecres
39	31	83.8	417	4 AAU63550	Aau63550 Propionib
40	31	83.8	417	6 ABM60069	Abm60069 Propionib
41	31	83.8	440	7 ABO74167	AbO74167 Pseudomon
42	31	83.8	448	6 ABP77283	Abp77283 N. gonorr
43	31	83.8	448	6 ABU37104	Abu37104 Protein e
44	31	83.8	475	7 ABH82627	Abh82627 Pseudomon
45	31	83.8	588	7 ADG20412	Adg20412 Pseudomon

#### ALIGNMENTS

RESULT 1	
AAB02893	
ID AAB02893 standard; peptide; 7 AA.	
XX	
AC AAB02893;	
XX	
DT 18-SEP-2000 (first entry)	
XX	
DE Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:34.	
XX	
KW Zinc finger domain; alpha helix; nucleotide binding; DNA binding;	
KW polylactyl protein; asymmetric target recognition;	
KW gene specific transcriptional regulator; gene activator; gene repressor;	
KW transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;	
KW transgenic animal; antiviral; anticancer; diagnosis.	
XX	
OS Synthetic.	
XX	
PN WO200023464-A2.	
XX	
PD 27-APR-2000.	
XX	
PF 14-OCT-1999; 99WO-EP007742.	
XX	
PR 16-OCT-1998; 98US-00173941.	
XX	
PA (NOVS ) NOVARTIS AG.	
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.	
PA (SCRI ) SCRIPPS RES INST.	
PI Barbas CF;	
XX	
DR WPI, 2000-339648/29.	
XX	
PT Novel isolated and purified zinc finger nucleotide-binding proteins with	
PT specificity for GNN triplet sequences, useful in gene therapy and for	
PT regulating gene function.	
XX	
PS Disclosure, Fig 1; 4bp; English.	
XX	
CC The invention relates to zinc finger nucleotide-binding proteins which	
CC comprise 2-12, preferably 2-6, operatively linked motifs selected from	
CC sequences AAB02860-802875. Sequences AAB02860-802875 represent the alpha	
CC helical regions of zinc finger domains which specifically bind to target	
CC nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked	
CC by the peptide linker TGRKP (AAB02970). The Cys2-His2 zinc finger motif	
CC is the most frequently utilised nucleic acid binding motif in eukaryotes,	
CC and constitutes a beta-beta-alpha fold. Nucleic acid recognition is	

Query Match 94.4%; Score 34; DB 2; Length 463;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DPSTLKR 7  
 |||||:  
 Db 52 DPSTLKR 58

## RESULT 2

094K10 PRELIMINARY; PRT; 527 AA.

AC 094K10; 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)  
 DE Mitochondrial processing peptidase.  
 OS Avicennia marina (Grey mangrove).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC Lamiales; Lamiaceae; Acanthaceae; Acanthaceae incertae sedis; Avicennia.  
 OC NCBI\_TaxID=82927;

RA Parani M., Parida A.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A363285; AAK51086.1; -.  
 DR HSSP; P10507; 1HR6.  
 DR MEROPS; M16.003; -.  
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
 DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR001431; Insulinase-like.  
 DR InterPro; IPR007863; Peptidase\_M16\_C.  
 DR Pfam; PF05193; Peptidase\_M16\_C; 1.  
 DR PROSITE; PS00143; INSULINASE; 1.

DR PROSITE; PS00143; INSULINASE; 1.

Query Match 94.4%; Score 34; DB 2; Length 527;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DPSTLKR 7  
 |||||:  
 Db 484 DPSTLKR 490

## RESULT 3

08P0G5 PRELIMINARY; PRT; 792 AA.

AC 08P0G5; 01-OCT-2002 (TReMBLrel. 22, Created)  
 DT 01-OCT-2002 (TReMBLrel. 22, last sequence update)  
 DT 01-MAR-2004 (TReMBLrel. 26, last annotation update)  
 DE Putative internalin A.  
 GN OrderedLocNames=SPY18\_1373;  
 OS Streptococcus pyogenes (serotype M18).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TaxID=186103;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

RA Parani M., Parida A.;

DR EMBL; AE010057; AAL97968.1; -.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR007092; LRR\_SD522.  
 DR InterPro; IPR006270; Strept\_his\_triad.  
 DR Pfam; PF00560; LRR\_1; 5.  
 DR Pfam; PF04270; Strept\_his\_triad; 3.  
 DR PRINTS; PR00019; LEURICHRPT.

DR PRINTS; PR00019; LEURICHRPT.

DR PRINTS; PR00019; LEURICHRPT.

Query Match 94.4%; Score 34; DB 2; Length 792;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DPSTLKR 7  
 |||||:  
 Db 299 DPSTLKR 305

## RESULT 4

099276 PRELIMINARY; PRT; 792 AA.

AC 099276; 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)  
 DT 01-MAR-2004 (TReMBLrel. 26, last annotation update)  
 DE Putative internalin A.  
 GN Name=InLA; OrderedLocNames=SPY1361;  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TaxID=1314;

RA Parani M., Parida A.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A363285; AAK51086.1; -.  
 DR HSSP; P10507; 1HR6.  
 DR MEROPS; M16.003; -.  
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.  
 DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR001431; Insulinase-like.  
 DR InterPro; IPR007863; Peptidase\_M16\_C.  
 DR Pfam; PF05193; Peptidase\_M16\_C; 1.  
 DR PROSITE; PS00143; INSULINASE; 1.

Query Match 94.4%; Score 34; DB 2; Length 792;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DPSTLKR 7  
 |||||:  
 Db 299 DPSTLKR 305

## RESULT 5

08K714 PRELIMINARY; PRT; 792 AA.

AC 08K714; 079XH7; 01-OCT-2002 (TReMBLrel. 22, Created)  
 DT 01-OCT-2002 (TReMBLrel. 22, last sequence update)  
 DT 25-OCT-2004 (TReMBLrel. 28, last annotation update)  
 DE Putative internalin A.  
 GN Name=InLA; OrderedLocNames=SPS0825, SPY13\_1035;  
 OS Streptococcus pyogenes (serotype M3).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.

Query Match 94.4%; Score 34; DB 2; Length 792;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds  
(without alignments)  
51.825 Million cell updates/sec

Title: US-10-646-919-33  
Perfect score: 36  
Sequence: 1 DPSTLKR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	94.4	463	2	Q6BR67 debaryomyce
2	34	94.4	527	2	Q94KI0 avicennia m
3	34	94.4	792	2	Q8P0G5 streptococc
4	34	94.4	792	2	Q99276 streptococc
5	34	94.4	792	2	Q8K714 streptococc
6	33	91.7	203	2	Q9L615 arbidopsia
7	33	91.7	204	2	Q79VU7 salmonella
8	33	91.7	204	2	Q924B8 plasmid col
9	33	91.7	204	2	Q7DJL6 plasmid r64
10	33	91.7	206	2	Q6XNAS rhodococcus
11	33	91.7	285	1	T2EB SCHPO
12	33	91.7	360	1	PLSK_DEIRA
13	33	91.7	363	2	Q8TVS1 delnoccoc
14	33	91.7	540	2	Q6MUD9 methanopyru
15	33	91.7	1316	2	Q7Y239 cryza sativ
16	33	91.7	1316	2	Q9SL02 arbidopsia
17	33	88.9	332	2	Q6BUB5 debaryomyce
18	32	88.9	375	2	Q7RVZ7 neurospora
19	32	88.9	447	1	Y4WB_RH1SN
20	32	88.9	577	2	Q92A03 listeria m
21	32	88.9	579	2	Q8Y5P5 listeria in
22	32	88.9	579	2	Q7IY10 listeria mo
23	32	88.9	654	2	Q9V5B3 dirosophila
24	32	88.9	1274	2	Q9UT05 schizosacch
25	32	88.9	1436	2	Q7R8K3 plasmodium
26	32	88.9	2313	2	Q782N1 neurospora
27	32	88.9	4246	2	Q7YU00 thermocoga
28	31	86.1	89	2	Q9XG08 giardia lam
29	31	86.1	113	2	Q7QT74 synchococc
30	31	86.1	162	2	Q8DX8 trichomonas
31	31	86.1	221	2	Q9Y175

32	31	86.1	224	2	Q988D9 rhizobium 1
33	31	86.1	251	1	Y136_PSESM
34	31	86.1	253	2	Q93T30 burkholderi
35	31	86.1	253	2	Q62ES9 burkholderi
36	31	86.1	253	2	Q63YB4 burkholderi
37	31	86.1	328	1	T2EB_YEAST
38	31	86.1	343	2	Q6CEB8 saccharomyc
39	31	86.1	372	1	AROB_PROMM
40	31	86.1	379	1	DXR_CHIMU
41	31	86.1	386	2	Q7MBW0 chlamydia m
42	31	86.1	386	2	Q8D526 vibrio vuln
43	31	86.1	393	2	Q97107 clostridium
44	31	86.1	422	2	Q6N329 rhodospseud
45	31	86.1	434	2	Q72R96 leptospira

## ALIGNMENTS

RESULT 1  
ID Q6BR67 PRELIMINARY; PRT; 463 AA.  
AC Q6BR67;  
DT 25-OCT-2004 (TREMBLrel. 28, Created)  
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)  
DE Similar to tr|Q93999 Candida albicans Possible zinc-finger protein.  
GN ORFNames=DEHA0D203179;  
OS Debaryomyces hansenii CBS767.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycetia; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
OX NCBI\_TaxID=284592;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CBS767;  
RG Genolevuree;  
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neugegisse C., Talla E., Goffard N., Frangeul L., Aigle M., Anthonard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C., Boissame A., Boyer J., Catolico L., Confanlet F., de Darvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Gropi A., Hantaye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerest A., Kozul R., Lemaire M., Lesur I., Ma L., Miller H., Nicard J.M., Nikolski M., Oxtas S., Olier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A., Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zenlou-Meyer M., Zivanovic I., Boliotin-Pukhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weisenbach J., Wincker P., Souciet J.L.;  
RT "Genome evolution in Yeasts."  
RL Nature 430:35-44(2004).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CBS767;  
RG Genoscope;  
RA Submitted (JUL-2004) to the EMBL/GenBank/DDA databases.  
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -1- SIMILARITY: Contains 1 Zn(2)-Cys(6) Fungal-type binuclear cluster domain.  
DR EMBL; CR382136; CAG87477.1; -;  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR001138; Fungi\_Tfscrp\_N.  
DR Pfam; PF00172; Zn\_cfus; 1.  
DR SMART; SM00066; GAL4; 1.  
DR PROSITE; PS50048; ZN2\_Cy6\_FUNGAL\_2; 1.  
KW DNA-binding; Metal-binding; Nuclear protein; Transcription; Transcription regulation; Zinc; Zinc-finger.  
SQ SEQUENCE 463 AA; 55028 MW; A73BBE827C2C0F73 CRC64;

Best Local Similarity 85.7%; Pred. No. 78;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DPSTLKR 7  
|||||:  
Db 169 DPSTLKR 175

RESULT 3  
T49600  
negative acting factor related protein [imported] - *Neurospora crassa*  
N:Alternate names: protein B3E4.130  
C:Species: *Neurospora crassa*  
C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49600  
R:Schulte, U.; Algu, V.; Hohnsels, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49600  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-375 <SCH>  
A:Cross-references: EMBL:AL355931; GSPDB:GN00116; NCSP:B3E4.130  
A:Experimental source: BAC clone B3E4; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B3E4.130  
A:Map position: 6  
A:Introns: 126/3

Query Match 88.9%; Score 32; DB 2; Length 375;  
Best Local Similarity 85.7%; Pred. No. 35;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DPSTLKR 7  
|||||:  
Db 30 DPSTLKR 36

RESULT 4  
AE1697  
two-component sensor histidine kinase homolog lln2119 [imported] - *Listeria innocua* (str  
C:Species: *Listeria innocua*  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AE1697  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A:Title: Comparative genomics of *Listeria species*.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AE1697  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-577 <GLA>  
A:Cross-references: UNIPROT:Q92A03; GB:AL592022; PIDN:CAC97349.1; PID:G16414633; GSPDB:G  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lln2119

Query Match 88.9%; Score 32; DB 2; Length 577;  
Best Local Similarity 85.7%; Pred. No. 55;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DPSTLKR 7  
|||||:  
Db 188 DPSTLKR 194

RESULT 5  
AC1326  
two-component sensor histidine kinase homolog lmo2011 [imported] - *Listeria monocytogenes*

C:Species: *Listeria monocytogenes*  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AC1326  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A:Title: Comparative genomics of *Listeria species*.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1326  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-579 <GLA>  
A:Cross-references: UNIPROT:Q8YSP5; GB:NC\_003210; PIDN:CAD00089.1; PID:G1641464; GSPDB:  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo2011

Query Match 88.9%; Score 32; DB 2; Length 579;  
Best Local Similarity 85.7%; Pred. No. 55;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DPSTLKR 7  
|||||:  
Db 190 DPSTLKR 196

RESULT 6  
T39249  
probable tripeptidylpeptidase - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T39249  
R:Medler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, September 1999  
A:Reference number: Z21742  
A:Accession: T39249  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1274 <MED>  
A:Cross-references: UNIPROT:Q9UT05; EMBL:AL117210; NID:el549906; PIDN:CAB55179.1; GSPDB:  
A:Experimental source: strain 972h-; clone pl p843  
C:Genetics:  
A:Gene: SPDB:SPAP8A3.12C  
A:Map position: 1

Query Match 88.9%; Score 32; DB 2; Length 1274;  
Best Local Similarity 85.7%; Pred. No. 13e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DPSTLKR 7  
|||||:  
Db 953 DPSTLKR 959

RESULT 7  
C72286  
transcription regulator, metal-sensing - *Thermotoga maritima* (strain MS8)  
C:Species: *Thermotoga maritima*  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: C72286  
R:Wilson, K.B.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C:M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: C72286  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-89 <ARN>

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds  
(without alignments)  
46.449 Million cell updates/sec

Title: US-10-646-919-33

Perfect score: 36

Sequence: 1 DPSTLKR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	91.7	285	2 T41052	transcription init
2	33	91.7	1292	2 D84727	probable RAD50 DNA
3	32	88.9	375	2 T49600	negative acting fa
4	32	88.9	577	2 AE1697	two-component sens
5	32	88.9	579	2 AC1326	probable tripeptid
6	32	88.9	1274	2 T39249	transcription regu
7	31	86.1	89	2 C72286	transcription init
8	31	86.1	328	2 S38138	1-deoxy-D-xylose
9	31	86.1	379	2 G81712	probable flavoprot
10	31	86.1	393	2 B97128	xanthine dehydroge
11	31	86.1	750	2 AE2860	50S ribosomal prot
12	31	86.1	779	2 D97637	ribosomal protein
13	30	83.3	44	2 AS2951	hypothetical prote
14	30	83.3	105	2 S52680	hypothetical prote
15	30	83.3	141	2 H90016	hypothetical prote
16	30	83.3	169	2 D84864	hypothetical prote
17	30	83.3	180	2 B70403	hypothetical prote
18	30	83.3	189	2 T15265	hypothetical prote
19	30	83.3	192	2 T21210	hypothetical prote
20	30	83.3	205	2 C89957	hypothetical prote
21	30	83.3	294	2 S75135	hypothetical prote
22	30	83.3	346	2 D90450	conserved hypotet
23	30	83.3	350	2 AD3560	methylnated-DNA-[pr
24	30	83.3	357	2 F83195	probable methylate
25	30	83.3	378	2 A97469	hypothetical prote
26	30	83.3	378	2 AE2587	hypothetical prote
27	30	83.3	407	2 AG0768	probable glycosyl
28	30	83.3	407	2 G90985	hypothetical prote
29	30	83.3	407	2 B85831	hypothetical prote

30	30	83.3	407	2 F55239	hypothetical 44.9K
31	30	83.3	429	2 S67101	probable microinat
32	30	83.3	462	2 T36848	probable glutamine
33	30	83.3	508	2 S07546	phocoystem II chl
34	30	83.3	610	2 F88109	protein T24E12.9 l
35	30	83.3	631	2 A54659	DNA repair protein
36	30	83.3	659	2 T02838	probable membrane
37	30	83.3	712	2 AD0274	probable toxin tra
38	30	83.3	775	2 F82887	hypothetical prote
39	30	83.3	1704	2 T43141	vitelllogenin 1 - m
40	30	83.3	3938	2 T42761	Basoon protein -
41	30	83.3	3942	2 T42730	Basoon protein -
42	29	80.6	77	2 U01295	hypothetical 8.8K
43	29	80.6	99	2 E71370	probable ribosoma
44	29	80.6	196	2 AF0715	probable ABC trans
45	29	80.6	201	1 MMWVBD	24K antigen - born

#### ALIGNMENTS

##### RESULT 1

T41052  
transcription initiation factor IIE beta subunit (TFIIE-BETA) - fission yeast (Schizosac  
C/Species: Schizosaccharomyces pombe  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #ext\_change 09-Jul-2004  
C/Accession: T41052  
R:Hilbert, H.; Duesterhoeft, A.; Wood, V.; Rajadream, M.A.; Barrel, B.G.  
submitted to the EMBL Data Library, August 1998  
A/Reference number: Z21967  
A/Accession: T41052  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-285 <HIL>  
A/Cross-references: UNIPROT:P79011; EMBL:AF031324; P1DN:CAA20446.1; GSPDB:GN00068; SPDB  
A/Experimental source: strain 97zh-; cosmid c1672  
C/Genetics:  
A/Gene: SPDB:SPCC1672.08c  
A/Map position: 3  
C/Keywords: transcription initiation

Query Match 91.7%; Score 33; DB 2; Length 285;  
Best Local Similarity 85.7%; Pred. No. 15;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7  
DB 244 DPSTVXR 250

##### RESULT 2

D84727  
probable RAD50 DNA repair protein [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #ext\_change 09-Jul-2004  
C/Accession: D84727  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.  
eues, D.; Nierman, W.C.; White, O.; Eilen, J.A.; Salzberg, S.L.; Fraeier, C.M.; Venter, J.  
Nature 407, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: D84727

A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1292 <STO>  
A/Cross-references: UNIPROT:Q9SL02; GB:AE002093; NID:G4263721; P1DN:AAD15407.1; GSPDB:G  
C/Genetics:  
A/Gene: At2g31970  
A/Map position: 2  
C/Superfamily: RAD50 protein  
Query Match 91.7%; Score 33; DB 2; Length 1292;



```
RESULT 2
US-10-474-792-158
; Sequence 158, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Louie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 158
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-158

Query Match
Best Local Similarity 94.4%; Score 34; DB 16; Length 343;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
DB 299 DPSTLKR 305

RESULT 3
US-10-870-507A-2
; Sequence 2, Application US/10870507A
; Publication No. US20040255351A1
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Shi, Jinrui
; TITLE OF INVENTION: Rad50 Polypeptides
; FILE REFERENCE: 1116ED
; CURRENT APPLICATION NUMBER: US/10/870,507A
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: 60/132,575
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: 09/538,396
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1316
; TYPE: PRT
; ORGANISM: Zea mays
US-10-870-507A-2

Query Match
Best Local Similarity 91.7%; Score 33; DB 16; Length 1316;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
DB 169 DPSTLKR 175

RESULT 4
US-10-437-963-161723
; Sequence 161723, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 161723
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Oryza sativa
; NAME/KEY: unsure
; LOCATION: (1)..(350)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_60881C.1.pep
US-10-437-963-161723

Query Match
Best Local Similarity 88.9%; Score 32; DB 16; Length 350;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
DB 316 DPSTLKR 322

RESULT 5
US-10-437-963-161726
; Sequence 161726, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Barbazuk, Brad
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 161726
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_60884C.1.pep
US-10-437-963-161726

Query Match
Best Local Similarity 88.9%; Score 32; DB 16; Length 351;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
DB 317 DPSTLKR 323

RESULT 6
US-10-425-115-303434
; Sequence 303434, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compen Gen Ltd.

OM protein - protein search, using SW model

Run on: July 20, 2005, 13:52:31 ; Search time 71.333 Seconds

(without alignments)  
38.093 Million cell updates/sec

Title: US-10-646-919-33

Perfect score: 36

Sequence: 1 DPSTLKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388188149 residues 1736639

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10D\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10E\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubppaa/US10F\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubppaa/US11\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	7	18 US-10-646-919-33	Sequence 33, Appl
2	34	94.4	343	16 US-10-474-792-158	Sequence 158, App
3	33	91.7	1316	16 US-10-870-507A-2	Sequence 2, Appl1
4	32	88.9	350	16 US-10-437-963-161723	Sequence 161723,
5	32	88.9	351	16 US-10-437-963-161726	Sequence 161726,
6	32	88.9	390	16 US-10-425-115-303424	Sequence 303424,
7	32	88.9	400	15 US-10-425-114-64694	Sequence 64694, A
8	32	88.9	1156	16 US-10-794-897-6	Sequence 6, Appl1
9	32	88.9	1165	13 US-10-026-188-2	Sequence 2, Appl1
10	32	88.9	1274	15 US-10-369-493-2308	Sequence 2308, Ap
11	31	86.1	87	16 US-10-425-115-279521	Sequence 279521,

12	31	86.1	126	15 US-10-424-599-206235	Sequence 206235,
13	31	86.1	141	16 US-10-437-963-140157	Sequence 140157,
14	31	86.1	187	15 US-10-424-599-227418	Sequence 227418,
15	31	86.1	205	14 US-10-032-585-7467	Sequence 7467, Ap
16	31	86.1	221	17 US-10-732-923-22523	Sequence 22523, A
17	31	86.1	233	15 US-10-425-114-41550	Sequence 41550, A
18	31	86.1	238	16 US-10-425-115-19186	Sequence 19186,
19	31	86.1	329	10 US-09-893-519A-6	Sequence 6, Appl1
20	31	86.1	335	13 US-10-062-254-356	Sequence 356, App
21	31	86.1	343	15 US-10-424-599-280346	Sequence 280346,
22	31	86.1	341	16 US-10-425-115-266539	Sequence 266539,
23	31	86.1	393	15 US-10-282-122A-51738	Sequence 51738, A
24	31	86.1	495	9 US-09-738-626-6583	Sequence 6583, Ap
25	31	86.1	1410	13 US-10-014-717-3	Sequence 3, Appl1
26	31	86.1	2245	17 US-10-732-923-18446	Sequence 18446, A
27	30	83.3	7	18 US-10-646-919-34	Sequence 34, Appl
28	30	83.3	4	9 US-09-815-242-12118	Sequence 12118, A
29	30	83.3	44	15 US-10-282-122A-66859	Sequence 66859, A
30	30	83.3	53	15 US-10-424-599-187972	Sequence 187972,
31	30	83.3	58	16 US-10-425-115-184765	Sequence 184765,
32	30	83.3	83	15 US-10-424-599-173274	Sequence 173274,
33	30	83.3	141	17 US-10-470-048B-221	Sequence 221, App
34	30	83.3	178	15 US-10-425-114-67865	Sequence 67865, A
35	30	83.3	181	16 US-10-767-701-51786	Sequence 51786, A
36	30	83.3	189	14 US-10-179-766-8	Sequence 8, Appl1
37	30	83.3	201	9 US-09-815-242-5400	Sequence 5400, Ap
38	30	83.3	205	9 US-09-815-242-12561	Sequence 12561, A
39	30	83.3	205	15 US-09-815-242-13029	Sequence 13029, A
40	30	83.3	205	15 US-10-282-122A-44173	Sequence 44173, A
41	30	83.3	205	17 US-10-857-625-802	Sequence 802, App
42	30	83.3	207	15 US-10-282-122A-71415	Sequence 71415, A
43	30	83.3	241	14 US-10-156-761-14781	Sequence 14781, A
44	30	83.3	257	15 US-10-296-115-1323	Sequence 1323, Ap
45	30	83.3	297	14 US-10-166-087-28	Sequence 28, Appl

## ALIGNMENTS

RESULT 1  
US-10-646-919-33  
; Sequence 33, Application US/10646919  
; Publication No. US20050148075A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbas, Carlos F.  
; TITLE OF INVENTION: Zinc-finger binding domains for GNN  
; FILE REFERENCE: TSRI 645.2C1  
; CURRENT FILING DATE: 2003-08-21  
; PRIOR FILING DATE: 2003-08-21  
; PRIOR APPLICATION NUMBER: US 09/494,190  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: PCT/EP99/07742  
; PRIOR FILING DATE: 1999-10-14  
; PRIOR APPLICATION NUMBER: US 09/173,941  
; PRIOR FILING DATE: 1998-10-16  
; NUMBER OF SEQ ID NOS: 129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 33  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthesized  
US-10-646-919-33

Query Match 100.0%; Score 36; DB 18; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7  
DB 1 DPSTLKR 7

FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:phage display  
OTHER INFORMATION: selected and mutagenized  
US-09-494-190-33

Query Match 100.0%; Score 36; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 4.1e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7  
1 DPSTLKR 7

RESULT 3  
US-09-645-835A-4  
Sequence 4, Application US/09645835A  
Patent No. 6833356  
GENERAL INFORMATION:  
APPLICANT: Heinrichs, Jon  
APPLICANT: Johnson, Scott  
APPLICANT: Koenig, Scott  
APPLICANT: Adamou, John E.  
TITLE OF INVENTION: Pneumococcal Protein Homologs and Fragments for  
TITLE OF INVENTION: Vaccines  
FILE REFERENCE: 469201-493  
CURRENT APPLICATION NUMBER: US/09/645,835A  
CURRENT FILING DATE: 2000-08-25  
PRIOR APPLICATION NUMBER: U.S. 60/150,750  
PRIOR FILING DATE: 1999-08-25  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 792  
TYPE: PRT  
ORGANISM: Streptococcus pyogenes  
US-09-645-835A-4

Query Match 94.4%; Score 34; DB 4; Length 792;  
Best Local Similarity 85.7%; Pred. No. 58;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7  
1 DPSTLKR 7

Db 299 DPSTLKR 305

RESULT 4  
US-09-538-396-2  
Sequence 2, Application US/09538396  
Patent No. 6815578  
GENERAL INFORMATION:  
APPLICANT: Mahajan, Pramod B.  
APPLICANT: Shi, Jinrui  
TITLE OF INVENTION: Maize Rad50 Orthologue and Uses Thereof  
FILE REFERENCE: 1116  
CURRENT APPLICATION NUMBER: US/09/538,396  
CURRENT FILING DATE: 2000-03-29  
EARLIER APPLICATION NUMBER: 60/132,575  
EARLIER FILING DATE: 1999-05-05  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 1316  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-538-396-2

Query Match 91.7%; Score 33; DB 4; Length 1316;  
Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7

Db 169 DPSTLKR 175

RESULT 5  
US-09-248-796A-15691  
Sequence 15691, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 15691  
LENGTH: 345  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-15691

Query Match 88.9%; Score 32; DB 4; Length 345;  
Best Local Similarity 85.7%; Pred. No. 62;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7  
55 DPSTLKR 61

RESULT 6  
US-09-248-796A-15869  
Sequence 15869, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 15869  
LENGTH: 210  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-15869

Query Match 86.1%; Score 31; DB 4; Length 210;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPSTLKR 7  
180 DPSTLKR 186

RESULT 7  
US-09-538-092-514  
Sequence 514, Application US/09538092  
Patent No. 6753314  
GENERAL INFORMATION:  
APPLICANT: Glot, Iolci  
APPLICANT: Manfield, Traci A.

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds  
(without alignments)  
24.117 Million cell updates/sec

Title: US-10-646-919-33

Perfect score: 36

Sequence: 1 DPSTLKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA:\*

1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

3: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

4: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

5: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

6: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

7: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

8: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

9: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

10: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

11: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

12: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

13: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

14: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

15: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

16: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

17: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

18: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

19: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

20: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

21: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

22: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

23: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

24: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

25: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

26: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*

27: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	7	3	US-09-173-941-33
2	36	100.0	7	4	US-09-494-190-33
3	34	94.4	7	4	US-09-645-835A-4
4	33	91.7	1316	4	US-09-538-396-2
5	32	88.9	345	4	US-09-248-796A-15691
6	31	86.1	210	4	US-09-248-796A-15869
7	31	86.1	328	4	US-09-538-092-514
8	31	86.1	1410	3	US-09-335-409-3
9	31	86.1	1410	3	US-09-568-102-3
10	31	86.1	1410	3	US-09-567-969-3
11	31	86.1	1410	3	US-09-568-480-3
12	31	86.1	1410	3	US-09-568-486-3
13	31	86.1	1410	3	US-09-568-472-3
14	31	86.1	1410	3	US-09-567-899-3
15	30	83.3	7	3	US-09-173-941-33
16	30	83.3	7	4	US-09-494-190-33
17	30	83.3	60	4	US-09-270-767-57387
18	30	83.3	72	4	US-09-252-991A-18994
19	30	83.3	105	4	US-09-538-092-153
20	30	83.3	323	4	US-09-270-767-42120
21	30	83.3	412	4	US-09-502-540-11614
22	30	83.3	571	4	US-09-902-540-14983
23	30	83.3	830	4	US-09-252-991A-22004
24	30	83.3	1341	4	US-09-248-796A-19074
25	29	80.6	108	4	US-09-540-236-3176
26	29	80.6	201	3	US-08-369-822C-4
27	29	80.6	201	3	US-08-779-764A-20

28	29	80.6	201	3	US-08-779-764A-21	Sequence 21, Appl
29	29	80.6	201	3	US-08-779-764A-22	Sequence 22, Appl
30	29	80.6	201	3	US-08-582-776C-4	Sequence 4, Appl
31	29	80.6	201	3	US-08-434-831B-4	Sequence 4, Appl
32	29	80.6	201	4	US-09-563-456-20	Sequence 20, Appl
33	29	80.6	201	4	US-09-563-456-21	Sequence 21, Appl
34	29	80.6	201	4	US-09-563-456-22	Sequence 22, Appl
35	29	80.6	219	4	US-09-543-681A-5362	Sequence 5362, Ap
36	29	80.6	221	4	US-09-252-991A-28805	Sequence 28805, A
37	29	80.6	222	4	US-09-252-991A-11510	Sequence 11510, A
38	29	80.6	222	4	US-09-270-767-41947	Sequence 41947, A
39	29	80.6	339	4	US-09-248-796A-15770	Sequence 15770, A
40	29	80.6	416	4	US-09-902-540-12512	Sequence 12512, A
41	29	80.6	696	4	US-09-134-000C-4874	Sequence 4874, Ap
42	29	80.6	1165	4	US-09-949-016-6874	Sequence 6874, Ap
43	29	80.6	1165	4	US-09-949-016-11392	Sequence 11392, A
44	28	77.8	18	6	5252328-14	Patent No. 5252328
45	28	77.8	18	6	5252328-14	Patent No. 5252328

#### ALIGNMENTS

```
RESULT 1
; US-09-173-941-33
; Sequence 33, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 33
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-33

Query Match      100.0%; Score 36; DB 3; Length 7;
Best local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DPSTLKR 7
      |||||
Db      1 DPSTLKR 7

RESULT 2
US-09-494-190-33
; Sequence 33, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 33
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

Db 299 DPSTIKR 305

# RESULT 6

ID ADR83938 standard; protein; 792 AA.

XX ADR83938;

XX 02-DEC-2004 (first entry)

XX S. pyogenes hyperimmune system reactive antigen SpY1361 DNA.

XX hyperimmune serum reactive antigen; vaccine; anticaline.

XX Streptococcus pyogenes.

XX WO2004078907-A2.

XX 16-SEP-2004.

XX 02-MAR-2004; 2004WO-EP02087.

XX 04-MAR-2003; 2003EP-00450061.

XX (INTE-) INTERCELL AG.

XX Meinke A, Nagy E, Winkler B, Gelbmann D;

XX MPI; 2004-653698/63.

XX N-PSDB; ADR83788.

PT New isolated nucleic acid molecules encoding hyperimmune serum-reactive antigens from Streptococcus pyogenes, useful for diagnosing, preventing and treating S. pyogenes infections.

XX Claim 14; SEQ ID NO 206; 145bp; English.

CC This invention describes a novel nucleic acid molecule encoding a hyperimmune serum reactive antigen or its fragment from Streptococcus pyogenes. The nucleic acid molecule or hyperimmune serum-reactive antigen or its fragment are useful for the manufacture of a pharmaceutical preparation, especially a vaccine, against S. pyogenes infection. In addition, the hyperimmune serum reactive antigen or fragment is used for the isolation and/or purification and/or identification of an interaction partner of the hyperimmune serum reactive antigen or its fragment, for the generation of a peptide (e.g. anticalines) binding to the antigen or fragment, or for the manufacture of a functional nucleic acid selected from aptamers and Spiegelmers. The nucleic acid molecule may also be used for the manufacture of functional ribonucleic acids, such as ribozymes, antisense nucleic acids and siRNA. ADR83733-ADR84189 represent S. pyogenes hyperimmune serum reactive antigens, fragments and the encoding polynucleotide described in the invention.

XX Sequence 792 AA;

Query Match 94.4%; Score 34; DB 8; Length 792;

Best Local Similarity 85.7%; Pred. No. 2.4e+02; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTIKR 7

Db 299 DPSTIKR 305

# RESULT 7

ID ADM26711 standard; protein; 363 AA.

XX ADM26711;

XX 20-MAY-2004 (first entry)

DE Hyperthermophile Methanopyrus kandleri protein #1317.

XX hyperthermophile; protein stability enhancement;

XX protein activity enhancement.

XX Methanopyrus kandleri.

XX WO2003076575-A2.

XX 18-SEP-2003.

XX 04-MAR-2003; 2003WO-US006664.

XX 04-MAR-2002; 2002US-0361742P.

XX 14-MAY-2002; 2002US-0380423P.

XX 16-SEP-2002; 2002US-0410974P.

XX (FIDE-) FIDELITY SYSTEMS INC.

XX (MALY/) MALYKH A.

XX Slesarev AI, Pavlov A, Pavlova N, Kozayavkin S;

XX MPI; 2003-748383/70.

XX N-PSDB; ADM27081.

XX Claim 31; SEQ ID NO 1317; 1023bp; English.

CC The invention comprises the amino acid sequence of proteins from the hyperthermophile Methanopyrus kandleri, the invention also comprises the complete genome from Methanopyrus kandleri. The Methanopyrus kandleri proteins of the invention are useful for enhancing the stability and/or activity of other proteins. The Methanopyrus kandleri genome is useful in a variety of diagnostic and analytical methods. The present amino acid sequence represents a Methanopyrus kandleri protein of the invention.

XX Sequence 363 AA;

Query Match 91.7%; Score 33; DB 7; Length 363;

Best Local Similarity 85.7%; Pred. No. 1.7e+02; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTIKR 7

Db 137 DPSTIKR 143

# RESULT 8

ID ABB77986 standard; protein; 1292 AA.

XX ABB77986;

XX 22-OCT-2002 (first entry)

XX Amino acid sequence of an Arabidopsis RAD50 homologue.

XX Nucleic acid integration; homologous recombination; telomeric region;

XX RAD50.

XX Arabidopsis thaliana.

XX EP1217074-A1.

XX 26-JUN-2002.

XX 22-DEC-2000; 2000EP-00204693.

XX 22-DEC-2000; 2000EP-00204693.

```
RESULT 4
AAU00027
ID AAU00027 standard; protein; 792 AA.
XX
XX AAU00027;
XX
XX 11-MAY-2001 (first entry)
XX
XX Streptococcus pyogenes GAS36(2).
XX
XX Group A streptococci protein 36(2); GAS36(2); immunogen; vaccine;
XX antibody; necrotizing fasciitis; scarlet fever; sepsis; impetigo;
XX bacterial meningitis; otitis media; community-acquired pneumonia.
XX
XX Streptococcus pyogenes.
XX
XX Key Location/Qualifiers
FH 94..99
FT Region /label= Histidine triad motif
FT /note= "Proposed to be involved in metal binding"
FT 164..169
FT /label= Histidine triad motif
FT /note= "Proposed to be involved in metal binding"
FT 246..251
FT /label= Histidine triad motif
FT /note= "Proposed to be involved in metal binding"
FT 329..334
FT /label= Histidine triad motif
FT /note= "Proposed to be involved in metal binding"
XX
XX MO200114421-A1.
XX
XX 01-MAR-2001.
XX
XX 25-AUG-2000; 2000WO-US023417.
XX
XX 25-AUG-1999; 99US-0150750P.
XX
XX (MED1-) MEDIMOTNE INC.
XX
XX Koenig S, Heinrichs J, Johnson LS, Adamou JE;
XX
XX WPI; 2001-211305/21.
XX
XX N-PSDB; AAS00037.
XX
XX New polypeptides obtained from group A or B streptococci, especially
XX Streptococcus aureus homologous to Sp36 protein of Streptococcus
XX pneumoniae useful as antibacterial vaccines.
XX
XX Claim 4; Fig 5b; 62pp; English.
XX
XX The sequence represents Streptococcus pyogenes Group A Streptococci
XX protein 36(2), GAS36(2). A recombinant cell producing GAS36, GAS36(2) or
XX GAS36 is useful as a vaccine for vaccinating an animal, preferably a
XX human against infection by a bacterial organism such as a streptococcal
XX or streptococcal bacteria, and for treating a disease caused by group A
XX streptococci, group B streptococci or Streptococcus aureus in an animal
XX preferably a human. Vaccines and antibodies against the proteins of the
XX invention are useful in prophylaxis and/or treatment of diseases such as
XX meningitis, otitis media, community-acquired pneumonia, and many diseases
XX of newborns. The proteins are also used as immunogens to stimulate the
XX production of antibodies for use in passive immunotherapy, for use as
XX diagnostic reagents and for use as reagents in other processes such as
XX affinity chromatography
XX
XX Sequence 792 AA;
XX
XX Query Match 94.4%; Score 34; DB 4; Length 792;
XX Best Local Similarity 85.7%; Pred. No. 2.4e+02;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DPSTLKR 7
DB 299 DPSTLKR 105
XX
XX RESULT 5
XX ABP27504
XX ID ABP27504 standard; protein; 792 AA.
XX
XX ABP27504;
XX
XX 02-JUL-2002 (first entry)
XX
XX Streptococcus polypeptide SEQ ID NO 4184.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus pyogenes.
XX
XX MO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
XX
XX 24-NOV-2000; 2000GB-00028727.
XX
XX 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Y Rosl, Grandi G, Fraser C;
XX Tettein H;
XX
XX WPI; 2002-352536/38.
XX
XX N-PSDB; ABN68135.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
XX disease caused by Streptococcus bacteria, such as meningitis, and for
XX detecting a compound that binds to the protein.
XX
XX Claim 1; Page 3575; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (51), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (1), nucleic acids encoding (1), ABN6604-ABN71526 and
XX antibodies that bind (1) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (1) are used to detect Streptococcus in a
XX biological sample. (1) is used to determine whether a compound binds to
XX (1). A composition comprising (1) or a nucleic acid encoding (1) may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (1) may be used to recombinantly produce (1) and may be
XX used in gene therapy. Antibodies to (1) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins
XX
XX Sequence 792 AA;
XX
XX Query Match 94.4%; Score 34; DB 5; Length 792;
XX Best Local Similarity 85.7%; Pred. No. 2.4e+02;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

CC achieved through specific contacts from side chains of amino acid  
CC residues in the alpha helix. Each zinc finger can recognise a sub-site of  
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains  
CC allows the recognition of extended contiguous asymmetric DNA sequences.  
CC For example, a synthetic polypeptide protein containing six zinc finger  
CC domains can recognise an 18 bp sequence, and such proteins are  
CC potentially highly gene-specific. The novel nucleotide-binding zinc  
CC finger proteins may therefore be used in the development of artificial  
CC gene-specific transcriptional regulators. Such transcriptional switches  
CC may be used to regulate the expression of oncogenes such as erbB-2,  
CC overexpression of which is involved in malignant transformation. The  
CC proteins are therefore useful in the treatment of cancers, and may also  
CC be used to activate genes involved in fighting diseases, and to treat  
CC viral infections by inhibiting the synthesis of viral gene products. They  
CC may be used in DNA-based diagnostic applications. The proteins may also  
CC be used in producing functional gene knockout or activation in  
CC heterozygous transgenic animals. Proteins of the invention can  
CC discriminate between sequences which have a single base difference. This  
CC is manifested in a >100-fold decrease in affinity for the variant  
CC sequence. Gene activation and repression can be achieved by targeting  
CC within the gene transcript, suggesting that information obtained from  
CC expressed sequence tags may be sufficient for the construction of gene  
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix  
CC phage library peptides disclosed in the invention  
XX

## SQ Sequence 7 AA:

Query Match 100.0%; Score 36; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7  
Db 1 DPSTLKR 7

## RESULT 2

ADP61982  
ID ADP61982 standard; peptide; 7 AA.

AC ADP61982;

DT 12-FEB-2004 (first entry)

XX zinc finger binding region #33.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

KW zinc finger binding region.

XX Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

PF 26-JUN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.

PR 14-OCT-1999; 99WO-EP007742.

PA (SCRI ) SCRIPPS RES INST.

PI Barbas CF;

XX MPI; 2003-800134/75.

PT Regulating expression of nucleotide sequence that contains sequence 5'-  
PT (GNN)n-3', comprises exposing nucleotide sequence to composition  
PT comprising isolated and purified zinc finger-nucleotide binding  
PT polypeptide.  
XX  
XX Disclosure; SEQ ID NO 33; 46pp; English.

CC The invention relates to an expression of nucleotide sequence that  
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by  
CC exposing nucleotide sequence to composition comprising isolated and  
CC purified zinc finger-nucleotide binding polypeptide that contains a  
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-  
CC Arg. The invention is used in the regulation of the expression of the  
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present  
CC sequence represents the amino acid sequence of a zinc finger binding  
CC region.  
XX

## SQ Sequence 7 AA:

Query Match 100.0%; Score 36; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7  
Db 1 DPSTLKR 7

## RESULT 3

ADJ98424  
ID ADJ98424 standard; peptide; 7 AA.

XX ADJ98424;

DT 06-MAY-2004 (first entry)

DE zinc finger DNA binding peptide #95.

XX library; multimeric DNA binding polypeptide;

KW zinc finger DNA binding peptide; gene expression silencing;

XX gene expression enhancement.

XX Unidentified.

PN WO2003066828-A2.

PD 14-AUG-2003.

PF 07-FEB-2003; 2003WO-US003705.

PR 07-FEB-2002; 2002US-0354981P.

PA (SCRI ) SCRIPPS RES INST.

PI Barbas CF, Blancafort P;

XX MPI; 2003-731499/69.

PT New zinc finger library of multimeric DNA binding polypeptides, useful  
PT for sterically occluding the binding site of a natural transcription  
PT factor, and enhancing or silencing target gene expression.

XX Disclosure; Fig 20; 64pp; English.

CC The invention comprises a library of multimeric DNA binding polypeptides  
CC (e.g. zinc finger DNA binding polypeptides). The zinc finger binding  
CC polypeptides of the invention are useful for sterically occluding the  
CC binding site of a natural transcription factor and enhancing or silencing  
CC target gene expression. The present amino acid sequence represents a zinc  
CC finger DNA binding peptide of the invention.  
XX

## SQ Sequence 7 AA:

Query Match 100.0%; Score 36; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7  
Db 1 DPSTLKR 7

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:14:30 ; Search time 82.5 Seconds  
(without alignments)  
32.816 Million cell updates/sec

Title: US-10-646-919-33

Perfect score: 36  
Sequence: 1 DPSTLKR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	36	100.0	7	3	AAB02892	Aab02892 Nucleotid
2	36	100.0	7	7	ADP61982	Adf61982 Zinc fing
3	36	100.0	7	7	ADJ98424	Adj98424 Zinc fing
4	34	94.4	792	4	AAU00027	Aau00027 Streptoco
5	34	94.4	792	5	ABP27504	Abp27504 Streptoco
6	34	94.4	792	8	ADR83938	Adr83938 S. pyogen
7	33	91.7	363	7	ADM26711	Adm26711 Hyperther
8	33	91.7	1292	5	ABR77986	Abb77986 Amino aci
9	33	91.7	1316	4	AA827248	Abb27248 Maize RAD
10	32	88.9	579	5	ABR48516	Abb48516 Listeria
11	32	88.9	651	4	ABR58403	Abb58403 Drosophil
12	32	88.9	651	8	ADS96540	Ad96540 Drosophil
13	32	88.9	676	8	ADM90820	Adm90820 Human pha
14	32	88.9	928	8	ADM90818	Adm90818 Human pha
15	32	88.9	1156	8	ADR87165	Adr87165 Rat betar
16	32	88.9	1164	5	ABR83853	Abb83853 Rat L-TTP
17	32	88.9	1274	8	ADN19655	Adn19655 Bacterial
18	32	88.9	1885	8	ADM90819	Adm90819 Human pha
19	31	86.1	205	5	ABP73630	Abp73630 Candida a
20	31	86.1	328	5	AAU82946	Aau82946 S. cerevi
21	31	86.1	393	6	ABU33814	Abu33814 Protein e
22	31	86.1	434	8	ADR28242	Adr28242 Leptospi
23	31	86.1	495	4	AAQ32829	Aaq32829 C glutam
24	31	86.1	1410	3	AAV58574	Adv58574 Sorangium
25	31	86.1	1576	8	ADQ97959	Adq97959 Mouse can

26	30	83.3	7	3	AAB02893	Aab02893 Nucleotid
27	30	83.3	7	7	ADP61983	Adf61983 Zinc fing
28	30	83.3	7	7	ADJ98425	Adj98425 Zinc fing
29	30	83.3	44	4	AAU36525	Aau36525 Pseudomon
30	30	83.3	44	6	ABU38935	Abu38935 Protein e
31	30	83.3	72	7	ABO70248	Ab070248 Pseudomon
32	30	83.3	105	6	ABR53589	Ab53589 Protein s
33	30	83.3	105	7	ADK64318	Adk64318 Diaseae t
34	30	83.3	141	6	ABU19003	Abj19003 Pathogen
35	30	83.3	141	6	ABM73460	Abm73460 Staphyloc
36	30	83.3	158	4	ABR69105	Abb69105 Drosophil
37	30	83.3	158	8	ADS96486	Ad96486 Drosophil
38	30	83.3	169	3	AAQ43188	Aaq43188 Arabidops
39	30	83.3	169	3	AAQ07957	Aaq07957 Arabidops
40	30	83.3	189	6	AAO19958	Aao19958 C elegans
41	30	83.3	201	4	AAU33904	Aau33904 Staphyloc
42	30	83.3	205	4	AAU37436	Aau37436 Staphyloc
43	30	83.3	205	4	AAU36968	Aau36968 Staphyloc
44	30	83.3	205	6	ABU16249	Abu16249 Protein e
45	30	83.3	205	6	ABM71226	Abm71226 Staphyloc

#### ALIGNMENTS

RESULT 1  
AAB02892  
ID AAB02892 standard; peptide; 7 AA.  
AC  
XX AAB02892;  
XX

18-SEP-2000 (first entry)

Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:33.

Zinc finger domain; alpha helix; nucleotide binding; DNA binding;

KW polylactyl protein; asymmetric target recognition;

KW gene specific transcriptional regulator; gene activator; gene repressor;

KW transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;

KW transgenic animal; antiviral; anticancer; diagnosis.

OS Synthetic.

PN WO200023464-A2.

PD 27-APR-2000.

PF 14-OCT-1999; 99WO-EP007742.

PR 16-OCT-1998; 98US-00173941.

PA (NOVS ) NOVARTIS AG.

PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI (SCRI ) SCRIPPS RES INST.

PI Barbas CF;

DR WPI, 2000-339648/29.

XX Novel isolated and purified zinc finger nucleotide-binding proteins with

PT specificity for GNN triplet sequences, useful in gene therapy and for

PT regulating gene function.

PS Disclosure, Fig 1; 48pp; English.

CC The invention relates to zinc finger nucleotide-binding proteins which

CC comprise 2-12, preferably 2-6, operatively linked motifs selected from

CC sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha

CC helical regions of zinc finger domains which specifically bind to target

CC nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked

CC by the peptide linker TGRKP (AAB02970). The Cys2-His2 zinc finger motif

CC is the most frequently utilised nucleic acid binding motif in eukaryotes,

CC and constitutes a beta-beta-alpha fold. Nucleic acid recognition is



DR HSP: P04050; 1150.  
DR GO: GO:0003677; F:DNA binding; IEA.  
DR GO: GO:0003899; F:DNA-directed RNA polymerase activity; IEA.  
DR GO: GO:0006350; P:transcription; IEA.  
DR InterPro: IPR007075; RNA\_pol\_Rpb1\_6.  
DR Pfam: PF04992; RNA\_pol\_Rpb1\_6; 1.  
FT NON\_TER 1 91  
FT NON\_TER 91 1  
SQ SEQUENCE 91 AA; 10475 MW; ACA2762E06ABBEC CRC64;  
Query Match 92.1%; Score 35; DB 2; Length 91;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCRLSR 7  
Db 20 GCRLSK 26  
RESULT 3  
O9BNP8 PRELIMINARY; PRT; 91 AA.  
ID O9BNP8;  
AC O9BNP8;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE RNA polymerase II largest subunit (Fragment).  
OS Bacthopolys multidentatus.  
OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Chilopoda;  
OC Pleurostigmophora; Lithobiomorpha; Ethropolyidae; Bacthopolys.  
ON NCBI\_Taxid=118455;  
RX [1]  
RP SEQUENCE FROM N.A.  
RA Regier J.C., Shultz J.W.;  
RT "A phylogenetic analysis of Myriapoda (Arthropoda) using two nuclear  
RT protein-encoding genes.";  
RL Zool. J. Linn. Soc. 132:469-486(2001).  
DR EMBL: AF240897; AAK11874.1; -.  
DR HSP: P04050; 1150.  
DR GO: GO:0003677; F:DNA binding; IEA.  
DR GO: GO:0003899; F:DNA-directed RNA polymerase activity; IEA.  
DR GO: GO:0006350; P:transcription; IEA.  
DR InterPro: IPR007075; RNA\_pol\_Rpb1\_6.  
DR Pfam: PF04992; RNA\_pol\_Rpb1\_6; 1.  
FT NON\_TER 1 91  
FT NON\_TER 91 1  
SQ SEQUENCE 91 AA; 10457 MW; ACA2736B12FFBEC CRC64;  
Query Match 92.1%; Score 35; DB 2; Length 91;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCRLSR 7  
Db 20 GCRLSK 26  
RESULT 4  
O6JSV7 PRELIMINARY; PRT; 108 AA.  
ID O6JSV7;  
AC O6JSV7;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE RNA polymerase II largest subunit (Fragment).  
OS Lithobius forficatus.  
OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Chilopoda;  
OC Pleurostigmophora; Lithobiomorpha; Lithobiidae; Lithobius.  
ON NCBI\_Taxid=245740;  
RX [1]  
RP SEQUENCE FROM N.A.  
RA Regier J.C.;  
RT Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY310213; AA077115.1; -.  
DR GO: GO:0003677; F:DNA binding; IEA.  
DR GO: GO:0003899; F:DNA-directed RNA polymerase activity; IEA.  
DR GO: GO:0006350; P:transcription; IEA.  
DR InterPro: IPR007081; RNA\_pol\_Rpb1\_5.  
DR InterPro: IPR007075; RNA\_pol\_Rpb1\_6.  
DR Pfam: PF04998; RNA\_pol\_Rpb1\_5; 1.  
DR Pfam: PF04992; RNA\_pol\_Rpb1\_6; 1.  
FT NON\_TER 1 108  
FT NON\_TER 108 1  
SQ SEQUENCE 108 AA; 12041 MW; D3BD188345BFPD9C CRC64;  
Query Match 92.1%; Score 35; DB 2; Length 108;  
Best Local Similarity 85.7%; Pred. No. 13;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCRLSR 7  
Db 20 GCRLSK 26  
RESULT 5  
O823V0 PRELIMINARY; PRT; 228 AA.  
ID O823V0;  
AC O823V0; Q7C7B2;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Probable global regulatory protein.  
GN Name=kgdr; OrderedlocusNames=STRY3251, t3010;  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
ON NCBI\_Taxid=601;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrall B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Ty2 / ATCC 700931;  
RC MEDLINE=22531367; PubMed=12644504;  
RX DOI=10.1128/JB.185.7.2330-2337.2003;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burdand V., Kodoyianni V., Schwartz D.C., Blattner F.R.,  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
RT and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
DR EMBL: AL627277; CAD02922.1; -.  
DR EMBL: AE016844; AA070562.1; -.  
KW Complete proteome.  
SQ SEQUENCE 228 AA; 25191 MW; 50EA8A3B87B9D07D CRC64;  
Query Match 92.1%; Score 35; DB 2; Length 228;  
Best Local Similarity 85.7%; Pred. No. 28;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCRLSR 7  
Db 27 GCRLSK 33

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds  
(without alignments)  
51.825 Million cell updates/sec

Title: US-10-646-919-32  
Perfect score: 38  
Sequence: 1 GCRELSR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues  
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	92.1	91	2 Q6JSY5	Q6JSY5 australoblu
2	35	92.1	91	2 Q9BNK6	Q9BNK6 pokabius bi
3	35	92.1	91	2 Q9BNP8	Q9BNP8 bothroplys
4	35	92.1	108	2 Q6JSV7	Q6JSV7 lithobius f
5	35	92.1	228	2 Q8Z3V0	Q8Z3V0 salmonella
6	35	92.1	228	2 Q8ZM49	Q8ZM49 salmonella
7	34	89.5	92	2 Q00385	Q00385 coccidioid
8	34	89.5	93	2 Q7S8T6	Q7S8T6 neuropept
9	33	86.8	140	2 Q6CYK9	Q6CYK9 erwinia car
10	33	86.8	216	2 Q700H1	Q700H1 usellago ma
11	33	86.8	216	2 Q700H8	Q700H8 usellago ma
12	33	86.8	216	2 Q700I0	Q700I0 usellago ma
13	33	86.8	248	2 Q7CP30	Q7CP30 agrobacteri
14	33	86.8	254	2 Q7DJ36	Q7DJ36 agrobacteri
15	33	86.8	254	2 Q9WMC0	Q9WMC0 agrobacteri
16	33	86.8	273	2 Q8VMB8	Q8VMB8 arabidopsi
17	33	86.8	410	1 B4_USTMA	B4_USTMA usellago ma
18	33	86.8	756	2 Q8DKX1	Q8DKX1 synchococc
19	33	86.8	767	2 Q7Q6S8	Q7Q6S8 anopheles g
20	32	84.2	132	2 Q26405	Q26405 brugia mala
21	32	84.2	151	2 Q9KGF1	Q9KGF1 bacillus ha
22	32	84.2	175	2 Q7XIG9	Q7XIG9 leprospiri
23	32	84.2	235	2 Q9GTI8	Q9GTI8 wucheria
24	32	84.2	273	2 Q7SDP4	Q7SDP4 ashbya goss
25	32	84.2	282	2 Q8TXS8	Q8TXS8 methanopyru
26	32	84.2	341	2 Q8ZVY1	Q8ZVY1 pyrobaculum
27	32	84.2	352	2 Q04963	Q04963 brugia paha
28	32	84.2	380	2 Q8VHGS	Q8VHGS mus musculu
29	32	84.2	386	2 Q6KZ33	Q6KZ33 picrophilus
30	32	84.2	539	2 Q9AS88	Q9AS88 oryza sativ
31	32	84.2	645	2 Q8VDM7	Q8VDM7 mus musculu

32	32	84.2	655	1 NA10_HUMAN	Q86w26 homo sapien
33	32	84.2	708	2 Q91YS3	Q91YS3 mus musculu
34	32	84.2	784	2 Q87ZN8	Q87ZN8 pseudomonas
35	32	84.2	846	2 Q8IXT0	Q8IXT0 homo sapien
36	32	84.2	853	2 Q8C1A7	Q8C1A7 mus musculu
37	32	84.2	1062	1 NA12_HUMAN	Q9nu02 homo sapien
38	32	84.2	1116	1 PLAS_HUMAN	Q9hau0 homo sapien
39	32	84.2	1205	2 Q6ZPK1	Q6ZPK1 mus musculu
40	32	84.2	1421	2 Q7XSL9	Q7XSL9 oryza sativ
41	32	84.2	2078	2 Q8WPN1	Q8WPN1 oikopleura
42	31	81.6	103	2 Q6XNB7	Q6XNB7 rhodococcus
43	31	81.6	309	1 DUSC_AZOVI	P96192 azotobacter
44	31	81.6	343	2 Q6DTJ9	Q6dt39 paenibacill
45	31	81.6	346	1 VANL_STRCO	Q9xak7 streptomyce

ALIGNMENTS

RESULT 1  
Q6JSY5 PRELIMINARY; PRT; 91 AA.  
ID Q6JSY5  
AC Q6JSY5; 05-JUL-2004 (TREMblrel. 27, Created)  
DT 05-JUL-2004 (TREMblrel. 27, Last sequence update)  
DE RNA polymerase II largeet subunit (Fragment).  
OS Australobius, Metazoa; Arthropoda; Myriapoda; Chilopoda;  
OC Pleurostigmophora; Lithodromorpha; Lithodidae; Australobius.  
RX NCBI\_Taxid=126927;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Regier J.C.;  
RL Submitted (MAY-2003) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AY310185; F:DNA binding; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR InterPro; IPR007075; RNA\_pol\_Rpbl\_6.  
DR Pfam; PF04992; RNA\_pol\_Rpbl\_6; 1.  
FT NON\_TER 1  
FT NON\_TER 91  
SQ SEQUENCE 91 AA; 10445 MW; 788C828412FFA37D CRC64;  
QY 1 GCRELSR 7  
Db 20 GCRELSK 26  
Query Match 92.1%; Score 35; DB 2; Length 91;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
RESULT 2  
Q9BNK6 PRELIMINARY; PRT; 91 AA.  
ID Q9BNK6  
AC Q9BNK6; 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
DE 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
DE RNA polymerase II largeet subunit (Fragment).  
OS pokabius bilabiatu.  
OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Chilopoda;  
OC Pleurostigmophora; Lithodromorpha; Lithodidae; Pokabius.  
RX NCBI\_Taxid=116484;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Regier J.C.; Shultz J.W.;  
RT "A phylogenetic analysis of Myriapoda (Arthropoda) using two nuclear  
protein-encoding genes."  
RL Zool. J. Linn. Soc. 132:469-486 (2001).  
DR EMBL; AF240939; AAK11916.1; -

A:Gene: atsc  
A:Genome: plasmid

## Query Match

Best Local Similarity 86.8%; Score 33; DB 2; Length 248;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRELSR 7  
|||:|  
DB 35 GCRDMSR 41

## RESULT 3

hypothetical protein al15016 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120  
C:Accession: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AH2432  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurihara, T.; Sasanoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH2432  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1273 <KIR>  
A:Cross-references: UNIPROT:Q8YMB8; GB:BA000019; PIDN:BA6715.1; PID:G17134154; GSPDB:G  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: al15016

Query Match  
Best Local Similarity 86.8%; Score 33; DB 2; Length 273;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCRELSR 7  
|||:|  
DB 243 GCRSLSR 249

## RESULT 4

C32696  
b4 protein - smut fungus (Ustilago maydis)

C:Species: Ustilago maydis (corn smut)  
C:Date: 22-Jun-1990 #sequence\_revision 28-Aug-1992 #text\_change 09-Jul-2004  
C:Accession: C32696  
R:Schulz, B.; Bannett, F.; Dahl, M.; Schlesinger, R.; Schaefer, W.; Martin, T.; Herskowitz, Cell 60, 295-306, 1990  
A:Title: The b alleles of U. maydis, whose combinations program pathogenic development.  
A:Reference number: A32696; MUID:90124638; PMID:1967554  
A:Accession: C32696  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-409 <SCH>  
A:Cross-references: UNIPROT:P22018; GB:M58556  
A:Note: the authors translated the codon GAA for residue 56 as Lys  
C:Superfamily: mating-type locus protein b1  
C:Keywords: DNA binding; nucleus

Query Match  
Best Local Similarity 86.8%; Score 33; DB 2; Length 409;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRELS 6  
|||:|  
DB 108 GCRELS 113

## RESULT 5

B84122  
hypothetical protein BH3778 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C:Accession: B84122

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiraoka, Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: B84122

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <STO>  
A:Cross-references: UNIPROT:Q9K6F1; GB:AP001520; GB:BA000004; NID:G10176401; PIDN:BA8074  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH3778

Query Match  
Best Local Similarity 84.2%; Score 32; DB 2; Length 151;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCRELSR 7  
|||:|  
DB 17 GCREMER 23

## RESULT 6

S25548  
surface-associated glycoprotein gp15/400 - nematode (Brugia pahangi) (fragment)

C:Species: Brugia pahangi  
C:Date: 20-Feb-1995 #sequence\_revision 23-Aug-1997 #text\_change 09-Jul-2004  
C:Accession: A49246; S25548  
R:Tweddie, S.; Paxton, W.A.; Ingram, L.; Maizels, R.M.; McReynolds, L.A.; Selkirk, M.B. Exp. Parasitol. 76, 156-164, 1993  
A:Title: Brugia pahangi and Brugia malayi: a surface-associated glycoprotein (gp15/400)  
A:Reference number: A49246; MUID:93202227; PMID:8454024  
A:Accession: A49246  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-351 <TW2>  
A:Cross-references: UNIPROT:Q04963; EMBL:X68190  
C:Keywords: glycoprotein

Query Match  
Best Local Similarity 84.2%; Score 32; DB 2; Length 351;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCRELSR 7  
|||:|  
DB 145 GCRELSR 151

## RESULT 7

T35363  
D-alanine-D-alanine ligase (EC 6.3.2.4) A SC66T3.06 [similarity] - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004  
C:Accession: T35363  
R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, June 1999  
A:Reference number: Z21576  
A:Accession: T35363  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-346 <MUR>  
A:Cross-references: UNIPROT:Q9XAK7; EMBL:AL079348; PIDN:CAB45462.1; GSPDB:GN00070; SCOE  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOE:SC66T3.06  
C:Superfamily: D-alanine-D-alanine ligase  
C:Keywords: ligase

Query Match  
Best Local Similarity 81.6%; Score 31; DB 2; Length 346;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds  
(without alignments)  
46.449 Million cell updates/sec

Title: US-10-646-919-32

Perfect score: 38

Sequence: 1 GCRLSR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	92.1	228	2 AB0878	probable global re
2	33	86.8	248	2 AH3178	short chain dehydr
3	33	86.8	273	2 AH2432	hypothetical prote
4	33	86.8	409	2 C32696	b4 protein - smut
5	32	84.2	151	2 B84122	hypothetical prote
6	32	84.2	351	2 S25548	surface-associated
7	31	81.6	346	2 T35363	D-alanine-D-alanin
8	31	81.6	407	2 D69316	mRNA 3'-end proces
9	31	81.6	489	2 S44609	hypothetical prote
10	31	81.6	558	2 G83049	DNA repair protein
11	31	81.6	575	2 S43128	beta-D-glucosidase
12	31	81.6	712	2 C71419	hypothetical prote
13	31	81.6	919	2 T05746	hypothetical prote
14	30	78.9	137	2 E64845	ycdV protein - Bsc
15	30	78.9	319	2 T33609	hypothetical prote
16	30	78.9	365	2 B69114	conserved hypothet
17	30	78.9	410	2 B32696	mating-type locus
18	30	78.9	420	2 I51666	Mel-1c receptor su
19	30	78.9	456	2 A31857	ribonuclease inhib
20	30	78.9	585	2 E85040	hypothetical prote
21	30	78.9	782	2 AE2262	phosphotibosylform
22	30	78.9	1607	2 T04583	TMV resistance pro
23	29	76.3	89	2 E97731	hypothetical prote
24	29	76.3	120	2 D95980	hypothetical prote
25	29	76.3	130	2 D82647	hypothetical prote
26	29	76.3	144	2 T08710	hypothetical prote
27	29	76.3	142	2 F96967	probable sigma fac
28	29	76.3	176	2 T05698	hypothetical prote
29	29	76.3	188	2 T34881	hypothetical prote

30	29	76.3	189	2 C86683	prophage pil prote
31	29	76.3	211	2 C95347	Protein [imported
32	29	76.3	270	2 S75051	lactose transport
33	29	76.3	315	2 E90435	conserved hypothet
34	29	76.3	353	2 A10220	probable sugar tra
35	29	76.3	354	2 B69284	sarcosine oxidase,
36	29	76.3	356	2 T41764	AcMMPV orf18 - Bom
37	29	76.3	361	2 A86393	TK7.4 protein - A
38	29	76.3	378	2 G02313	CDG37 homolog - hu
39	29	76.3	379	2 T20394	hypothetical prote
40	29	76.3	402	2 C69110	glutamate N-acetyl
41	29	76.3	417	2 A97063	NAD(P)H-dependent
42	29	76.3	476	2 B84379	methylaspartate mu
43	29	76.3	496	2 T08674	probable finger pr
44	29	76.3	527	2 T37895	hypothetical prote
45	29	76.3	691	2 T32748	hypothetical prote

## ALIGNMENTS

RESULT 1  
AB0878  
probable global regulatory protein kdgr [imported] - Salmonella enterica subsp. enterica  
C:Species: Salmonella enterica subsp. enterica serovar Typh  
A:Note: this species has also been called Salmonella typh  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AB0878  
R:Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher,  
Th, T., Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar,  
, S., Moule, S., O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K.,  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AB0878  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-228 <PAR->  
A:Cross-references: GB:AL513382; PIDN:CAN02922.1; PID:gt16504175; GSPDB:GN00176  
C:Genetics:  
A:gene: kdgr

Query Match  
Best Local Similarity 92.1%; Score 35; DB 2; Length 228;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRLSR 7  
DB 27 GCRLSR 33

RESULT 2  
AH3178  
short chain dehydrogenase dehydrogenases atsc [imported] - Agrobacterium tumefaciens (st  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
C:Accession: AH3178  
R:Wood, D.W., Setudal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L  
erage, G., Gillet, W., Grant, C., Guenther, D., Kutyavhin, T., Levy, R., Li, M., McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H., Teo, Y., Biddle, P., Jung, M., Kreppan, W., Perry, M., Gordon-Kamm,  
ster, B.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AH3178  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-248 <KUR->  
A:Cross-references: UNIPROT:Q9WMC0; GB:AE008687; PIDN:AAL4846.1; PID:gt17743587; GSPDB:G  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:

```
RESULT 2
US-10-767-701-49350
; Sequence 49350, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 49350
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3476-048-P1-K1-A4.pep
US-10-767-701-49350

Query Match          92.1%; Score 35; DB 16; Length 66;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRCRSR 7
DB 40 GCRCRSR 46

RESULT 3
US-10-029-386-31910
; Sequence 31910, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31910
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004495.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.61
US-10-029-386-31910

Query Match          86.8%; Score 33; DB 14; Length 33;
Best Local Similarity 85.7%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCRCRSR 7
DB 26 GCRCRSR 32

RESULT 4
US-10-425-115-269178
; Sequence 269178, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 269178
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_177089C.1.pep
US-10-425-115-269178

Query Match          86.8%; Score 33; DB 16; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRCRS 6
DB 6 GCRCRS 11

RESULT 5
US-10-425-115-258881
; Sequence 258881, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 258881
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_167686C.1.pep
US-10-425-115-258881

Query Match          86.8%; Score 33; DB 16; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRCRS 6
DB 84 GCRCRS 89

RESULT 6
US-10-106-698-4323
; Sequence 4323, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruden et al.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:52:31 ; Search time 71.333 Seconds  
(without alignments)  
38.093 Million cell updates/sec

Title: US-10-646-919-32

Perfect score: 38

Sequence: 1 GCRESLR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	7	18	US-10-646-919-32
2	35	92.1	66	16	US-10-767-701-48350
3	33	86.8	33	14	US-10-029-386-31910
4	33	86.8	76	16	US-10-425-115-269178
5	33	86.8	90	16	US-10-425-115-269178
6	32	84.2	103	14	US-10-106-698-432881
7	32	84.2	131	15	US-10-437-963-145479
8	32	84.2	147	16	US-10-437-963-145479
9	32	84.2	173	16	US-10-767-701-39771
10	32	84.2	199	16	US-10-425-115-286717
11	32	84.2	317	15	US-10-407-866-113

12	32	84.2	412	15	US-10-094-749-2350	Sequence 2350, Ap
13	32	84.2	418	16	US-10-425-115-264314	Sequence 264314,
14	32	84.2	422	15	US-10-425-115-56584	Sequence 56584, A
15	32	84.2	504	10	US-09-965-621-22	Sequence 22, Appl
16	32	84.2	504	15	US-10-407-866-22	Sequence 22, Appl
17	32	84.2	504	16	US-10-781-294-22	Sequence 22, Appl
18	32	84.2	555	18	US-10-499-353A-655	Sequence 655, App
19	32	84.2	567	16	US-10-437-963-106841	Sequence 106841,
20	32	84.2	590	15	US-10-407-866-94	Sequence 94, Appl
21	32	84.2	612	16	US-10-437-963-186933	Sequence 186933,
22	32	84.2	650	15	US-10-407-866-85	Sequence 85, Appl
23	32	84.2	650	15	US-10-407-866-86	Sequence 86, Appl
24	32	84.2	655	14	US-10-124-498-13	Sequence 13, Appl
25	32	84.2	655	14	US-10-066-521-13	Sequence 13, Appl
26	32	84.2	655	14	US-10-296-539-2	Sequence 2, Appl1
27	32	84.2	655	15	US-10-407-866-84	Sequence 84, Appl
28	32	84.2	655	16	US-10-794-342-15	Sequence 15, Appl
29	32	84.2	770	16	US-10-343-663A-68	Sequence 68, Appl
30	32	84.2	1033	13	US-10-127-516-2	Sequence 2, Appl1
31	32	84.2	1033	13	US-10-027-629-2	Sequence 2, Appl1
32	32	84.2	1033	14	US-10-029-347-4	Sequence 4, Appl1
33	32	84.2	1033	14	US-10-132-967-2	Sequence 2, Appl1
34	32	84.2	1033	16	US-10-882-761-4	Sequence 4, Appl1
35	32	84.2	1062	14	US-10-239-663-43	Sequence 43, Appl
36	32	84.2	1094	15	US-10-275-107-60	Sequence 60, Appl
37	32	84.2	1116	9	US-09-790-318-2	Sequence 2, Appl1
38	32	84.2	1116	15	US-10-112-944-323	Sequence 323, App
39	32	84.2	1116	16	US-10-343-663A-27	Sequence 27, Appl
40	32	84.2	1421	16	US-10-437-963-182652	Sequence 182652,
41	31	81.6	43	16	US-10-425-115-25124	Sequence 215124,
42	31	81.6	70	15	US-10-424-599-149899	Sequence 149899,
43	31	81.6	70	16	US-10-425-115-313580	Sequence 313580,
44	31	81.6	82	16	US-10-425-115-283149	Sequence 283149,
45	31	81.6	104	16	US-10-425-115-196784	Sequence 196784,

#### ALIGNMENTS

RESULT 1  
US-10-646-919-32  
; Sequence 32, Application US/10646919  
; Publication No. US20050148075A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbas, Carlos F.  
; TITLE OF INVENTION: Zinc-finger binding domains for GNN  
; FILE REFERENCE: TSRI 645.2C1  
; CURRENT APPLICATION NUMBER: US/10/646,919  
; PRIOR FILING DATE: 2003-08-21  
; PRIOR APPLICATION NUMBER: US 09/494,190  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: PCT/EP99/07742  
; PRIOR FILING DATE: 1999-10-14  
; PRIOR APPLICATION NUMBER: US 09/173,941  
; PRIOR FILING DATE: 1998-10-16  
; NUMBER OF SEQ ID NOS: 129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthesized  
US-10-646-919-32

Query Match 100.0%; Score 38; DB 18; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.6e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRESLR 7  
DB 1 GCRESLR 7

```
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
; OTHER INFORMATION: selected and mutagenized
US-09-494-190-32

Query Match
Best Local Similarity 100.0%; Score 38; DB 4; Length 7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRELSR 7
   |||||
Db 1 GCRELSR 7

RESULT 3
US-09-270-767-33518
; Sequence 33518, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33518
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33518

Query Match
Best Local Similarity 81.6%; Score 31; DB 4; Length 153;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRELSR 7
   |||||
Db 84 GCKELSQ 90

RESULT 4
US-09-270-767-48735
; Sequence 48735, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48735
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-48735

Query Match
Best Local Similarity 81.6%; Score 31; DB 4; Length 153;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRELSR 7
   |||||
Db 84 GCKELSQ 90

RESULT 5
US-09-252-991A-30735
; Sequence 30735, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
```

```
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30735
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30735

Query Match
Best Local Similarity 81.6%; Score 31; DB 4; Length 425;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCRELSR 7
   |||||
Db 287 GCRDLGR 293

RESULT 6
US-09-252-991A-27035
; Sequence 27035, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27035
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27035

Query Match
Best Local Similarity 81.6%; Score 31; DB 4; Length 708;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRELSR 7
   |||||
Db 301 GCOELAR 307

RESULT 7
US-09-621-976-7590
; Sequence 7590, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dunas Milne Edwards, J.B.
; APPLICANT: Ubert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent. pm
; SEQ ID NO 7590
; LENGTH: 60
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds  
(without alignments)  
24.117 Million cell updates/sec

Title: US-10-646-919-32

Perfect score: 38  
Sequence: 1 GCRELSR 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A-COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B-COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCITUS-COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	7	3 US-09-173-941-32	Sequence 32, Appl
2	38	100.0	7	4 US-09-494-190-32	Sequence 32, Appl
3	31	81.6	153	4 US-09-270-767-33518	Sequence 33518, A
4	31	81.6	153	4 US-09-270-767-48735	Sequence 48735, A
5	31	81.6	425	4 US-09-252-991A-30735	Sequence 30735, A
6	31	81.6	708	4 US-09-252-991A-27035	Sequence 27035, A
7	30	78.9	60	4 US-09-621-976-7590	Sequence 7590, Ap
8	30	78.9	177	4 US-09-270-767-46995	Sequence 46995, A
9	30	78.9	247	4 US-10-101-464A-730	Sequence 23888, A
10	30	78.9	303	4 US-09-252-991A-23888	Sequence 23888, A
11	30	78.9	397	4 US-09-489-039A-9401	Sequence 9401, Ap
12	30	78.9	412	4 US-09-252-991A-26284	Sequence 26284, A
13	30	78.9	420	2 US-08-466-103A-2	Sequence 2, Appl1
14	30	78.9	456	2 US-08-910-731-2	Sequence 2, Appl1
15	30	78.9	456	2 US-08-795-395-2	Sequence 51, Appl
16	30	78.9	503	4 US-09-823-038A-51	Sequence 42310, A
17	30	78.9	648	4 US-09-270-767-42310	Sequence 7729, Ap
18	29	76.3	73	4 US-09-489-039A-7729	Sequence 5991, Ap
19	29	76.3	87	4 US-09-513-999C-5991	Sequence 5990, Ap
20	29	76.3	120	4 US-09-513-999C-5990	Sequence 23882, A
21	29	76.3	143	4 US-09-248-796A-23882	Sequence 46, Appl
22	29	76.3	143	1 US-08-253-155A-46	Sequence 11698, A
23	29	76.3	145	4 US-09-489-039A-11698	Sequence 20088, A
24	29	76.3	179	4 US-09-252-991A-20088	Sequence 424, App
25	29	76.3	193	4 US-09-205-258-424	Sequence 47629, A
26	29	76.3	193	4 US-09-270-767-47629	Sequence 23253, A
27	29	76.3	201	4 US-09-252-991A-23253	

28	29	76.3	206	1 US-08-468-847B-2	Sequence 2, Appl1
29	29	76.3	206	1 US-08-468-847B-20	Sequence 20, Appl
30	29	76.3	206	3 US-09-188-930-159	Sequence 159, App
31	29	76.3	206	3 US-09-188-930-266	Sequence 286, App
32	29	76.3	206	4 US-09-312-283C-159	Sequence 159, App
33	29	76.3	206	4 US-09-312-283C-286	Sequence 286, App
34	29	76.3	208	4 US-09-902-540-11472	Sequence 11472, A
35	29	76.3	212	4 US-09-902-540-12720	Sequence 12720, A
36	29	76.3	218	3 US-08-675-885-5	Sequence 5, Appl1
37	29	76.3	223	4 US-09-684-405-13	Sequence 13, Appl1
38	29	76.3	244	4 US-09-902-540-10220	Sequence 10220, A
39	29	76.3	246	4 US-09-252-991A-23684	Sequence 23684, A
40	29	76.3	271	4 US-09-540-236-2146	Sequence 2146, Ap
41	29	76.3	279	4 US-09-252-991A-24954	Sequence 24954, A
42	29	76.3	281	4 US-09-252-991A-23775	Sequence 23775, A
43	29	76.3	293	4 US-10-237-551-211	Sequence 211, App
44	29	76.3	310	4 US-09-543-681A-5247	Sequence 5247, Ap
45	29	76.3	313	4 US-09-252-991A-25230	Sequence 25230, A

## ALIGNMENTS

```

RESULT 1
; US-09-173-941-32
; Sequence 32, Application US/09173941
; Patent No. 614081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-32

Query Match      100.0%; Score 38; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;

QY      1 GCRELSR 7
|||||
Db      1 GCRELSR 7

RESULT 2
US-09-494-190-32
; Sequence 32, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence

```



QY 1 GCRELSR 7  
 DB 52 GCRELDR 58

RESULT 6  
 ID AAM06550 standard; protein; 74 AA.

AC AAM06550;

DT 05-OCT-2001 (first entry)

DE Human foetal protein, SEQ ID NO: 281.

XX Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;  
 KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;  
 KW gene therapy; antisense therapy; cancer; immune disorder;  
 KW growth disorder; osteoporosis; thrombolytic disorder;  
 KW nervous system disorder; inflammation.

OS Homo sapiens.

PN MO200155339-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US002723.

PR 25-JAN-2000; 2000US-00491404.

PR 15-SEP-2000; 2000US-00663870.

PR 06-NOV-2000; 2000US-00707351.

PA (HYSE-) HYSEQ INC.

PI Yeung G, Ford JB, Boyle BJ, Arterburn MC, Dmanac RA, Tang YT;

PI Liu C, Asundi V, Zhou P, Werhman T;

DR MPI; 2001-465571/50.

DR N-PSDB; AAH94225.

PT Novel fetal proteins useful for the treatment and diagnosis of diseases

PT associated with dysfunction of the protein e.g. cancers, immune

PT disorders, growth disorders, thrombolytic disorders, nervous system

PT disorders and inflammation.

CC Claim 10; Page 282-283; 715pp; English.

CC The invention relates to novel foetal polypeptides encoded by

CC polynucleotides comprising one of 477 sequences fully defined in the

CC specification. The foetal polynucleotides and polypeptides are useful in

CC the treatment and diagnosis of diseases such as cancers, immune

CC disorders, growth disorders (e.g. osteoporosis), thrombolytic disorders,

CC nervous system disorders and inflammation. The present sequence is a

CC polypeptide encoded by a cDNA assembled using an expressed sequence tag

CC (EST) found to be expressed in human foetal tissue cDNA libraries

XX Sequence 74 AA;

QY Query Match 89.5%; Score 34; DB 4; Length 74;

Best Local Similarity 85.7%; Pred. No. 64;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 GCRELSR 7

44 GCREFSR 50

RESULT 7

AAM06727

ID AAM06727 standard; protein; 74 AA.

XX

AC AAM06727;  
 DT 05-OCT-2001 (first entry)  
 DE Human foetal protein, SEQ ID NO: 458.

XX Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;  
 KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;  
 KW gene therapy; antisense therapy; cancer; immune disorder;  
 KW growth disorder; osteoporosis; thrombolytic disorder;  
 KW nervous system disorder; inflammation.

OS Homo sapiens.

PN MO200155339-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US002723.

PR 25-JAN-2000; 2000US-00491404.

PR 15-SEP-2000; 2000US-00663870.

PR 06-NOV-2000; 2000US-00707351.

PA (HYSE-) HYSEQ INC.

PI Yeung G, Ford JB, Boyle BJ, Arterburn MC, Dmanac RA, Tang YT;

PI Liu C, Asundi V, Zhou P, Werhman T;

DR MPI; 2001-465571/50.

DR N-PSDB; AAH94402.

PT Novel fetal proteins useful for the treatment and diagnosis of diseases

PT associated with dysfunction of the protein e.g. cancers, immune

PT disorders, growth disorders, thrombolytic disorders, nervous system

PT disorders and inflammation.

CC Claim 10; Page 345; 715pp; English.

CC The invention relates to novel foetal polypeptides encoded by

CC polynucleotides comprising one of 477 sequences fully defined in the

CC specification. The foetal polynucleotides and polypeptides are useful in

CC the treatment and diagnosis of diseases such as cancers, immune

CC disorders, growth disorders (e.g. osteoporosis), thrombolytic disorders,

CC nervous system disorders and inflammation. The present sequence is a

CC polypeptide encoded by a cDNA assembled using an expressed sequence tag

CC (EST) found to be expressed in human foetal tissue cDNA libraries

XX Sequence 74 AA;

QY Query Match 89.5%; Score 34; DB 4; Length 74;

Best Local Similarity 85.7%; Pred. No. 64;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 GCRELSR 7

44 GCREFSR 50

RESULT 8

AAM06747

ID AAM06747 standard; protein; 74 AA.

XX

AC AAM06747;

DT 05-OCT-2001 (first entry)

DE Human foetal protein, SEQ ID NO: 955.

XX Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;

KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;

KW gene therapy; antisense therapy; cancer; immune disorder;

KW growth disorder; osteoporosis; thrombolytic disorder;

```
RESULT 4
AAU53152
ID AAU53152 standard; protein; 65 AA.
XX
AC AAU53152;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #14048.
XX
KM SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-019047P.
XX
PR 02-JUN-2000; 2000US-0208841P.
XX
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR MPI: 2001-616774/71.
XX
DR N-PSDB; AAS59558.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX
Example 1; SEQ ID NO 14347; 1069pp; English.
XX
PS Sequences AAU3105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 65 AA;
XX
Query Match 89.5%; Score 34; DB 4; Length 65;
Best Local Similarity 85.7%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
RESULT 5
ABM49671
ID ABM49671 standard; protein; 65 AA.
XX
AC ABM49671;
XX
DT 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes predicted ORF-encoded polypeptide #14347.
XX
KM Acne vulgaris; antiseborrheic; dermatological; antibacterial;
KM immunostimulant; immune response; vaccine.
XX
OS Propionibacterium acnes.
XX
PN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieve-Douglass J;
XX
DR MPI: 2003-381789/36.
XX
DR N-PSDB; ACF64487.
XX
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
XX
Example 1; SEQ ID NO 14347; 1481pp; English.
XX
PS The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridization. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 65 AA;
XX
Query Match 89.5%; Score 34; DB 6; Length 65;
Best Local Similarity 85.7%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

CC achieved through specific contacts from side chains of amino acid  
CC residues in the alpha helix. Each zinc finger can recognise a subsite of  
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains  
CC allows the recognition of extended contiguous asymmetric DNA sequences.  
CC For example, a synthetic polypeptide protein containing six zinc finger  
CC domains can recognise an 18 bp sequence, and such proteins are  
CC potentially highly gene-specific. The novel nucleotide-binding zinc  
CC finger proteins may therefore be used in the development of artificial  
CC gene-specific transcriptional regulators. Such transcriptional switches  
CC may be used to regulate the expression of oncogenes such as *erbB-2*,  
CC overexpression of which is involved in malignant transformation. The  
CC proteins are therefore useful in the treatment of cancers, and may also  
CC be used to activate genes involved in fighting diseases, and to treat  
CC viral infections by inhibiting the synthesis of viral gene products. They  
CC may be used in DNA-based diagnostic applications. The proteins may also  
CC be used in producing functional gene knockout or activation in  
CC heterozygous transgenic animals. Proteins of the invention can  
CC discriminate between sequences which have a single base difference. This  
CC is manifested in a >100-fold decrease in affinity for the variant  
CC sequence. Gene activation and repression can be achieved by targeting  
CC within the gene transcript, suggesting that information obtained from  
CC expressed sequence tags may be sufficient for the construction of gene  
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix  
CC phage library peptides disclosed in the invention  
XX

SO Sequence 7 AA;

Query Match 100.0%; Score 38; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCELRSR 7  
| | | | |  
DB 1 GCCELRSR 7

RESULT 2  
ADP61981  
ID ADP61981 standard; peptide; 7 AA.

AC ADF61981;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Zinc finger binding region #32.  
XX  
XX zinc finger-nucleotide binding polypeptide; expression regulation;  
KM zinc finger binding region.  
XX

OS Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

PF 28-JAN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.

PR 14-OCT-1999; 99WO-EP007742.

PA (SCRI ) SCRIPPS RES INST.

XX Barbas CF;

PI Barbas CF;

XX WPI; 2003-800134/75.

PT Regulating expression of nucleotide sequence that contains sequence 5'-  
PT (GNN)n-3', comprises exposing nucleotide sequence to composition  
PT comprising isolated and purified zinc finger-nucleotide binding  
PT polypeptide.  
XX  
XX Claim 1; SEQ ID NO 32; 46pp; English.  
PS  
XX

CC The invention relates to an expression of nucleotide sequence that  
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by  
CC exposing nucleotide sequence to composition comprising isolated and  
CC purified zinc finger-nucleotide binding polypeptide that contains a  
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-  
CC Arg. The invention is used in the regulation of the expression of the  
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present  
CC sequence represents the amino acid sequence of a zinc finger binding  
CC region.  
XX

SO Sequence 7 AA;

Query Match 100.0%; Score 38; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCELRSR 7  
| | | | |  
DB 1 GCCELRSR 7

RESULT 3  
ADJ98423  
ID ADJ98423 standard; peptide; 7 AA.

AC ADJ98423;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Zinc finger DNA binding peptide #94.  
XX

KM library; multimeric DNA binding polypeptide;  
KM zinc finger DNA binding peptide; gene expression silencing;  
KW gene expression enhancement.  
XX

OS Unidentified.

PN WO2003066828-A2.

PD 14-AUG-2003.

PF 07-FEB-2003; 2003WO-US003705.

PR 07-FEB-2002; 2002US-0354981P.

XX (SCRI ) SCRIPPS RES INST.

XX Barbas CF, Blancafort P;

PI WPI; 2003-731499/69.

PT New zinc finger library of multimeric DNA binding polypeptides, useful  
PT for sterically occluding the binding site of a natural transcription  
PT factor, and enhancing or silencing target gene expression.  
XX

PS Disclosure; Fig 20; 64pp; English.

XX The invention comprises a library of multimeric DNA binding polypeptides  
XX (e.g. zinc finger DNA binding polypeptides). The zinc finger binding  
XX polypeptides of the invention are useful for sterically occluding the  
XX binding site of a natural transcription factor and enhancing or silencing  
XX target gene expression. The present amino acid sequence represents a zinc  
XX finger DNA binding peptide of the invention.  
XX

SO Sequence 7 AA;

Query Match 100.0%; Score 38; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCELRSR 7  
| | | | |  
DB 1 GCCELRSR 7

